RADIO ENGINEERING
and
ELECTRONIC PHYSICS

English Edition of

РАДИОТЕХНИКА
И
ЭЛЕКТРОНИКА

Published by the American Institute of Electrical Engineers
with the aid of a grant from the National Science Foundation

No.9 September 1961

Translated and Produced by Royer and Roger, Inc.

AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS

Established 1884

345 East Forty-Seventh Street
New York 17, N.Y.

Warren H. Chase, President
N.S. Hibshman, Executive Secretary
C.E. Dean, Technical Vice President, Communications
W.F. Denkhaus, Director of Publications
L.G. Abraham, Chairman, Communications Division

The English edition of RADIO ENGINEERING AND ELECTRONIC PHYSICS is published by the American Institute of Electrical Engineers with the aid of a grant from the National Science Foundation. © 1961 by American Institute of Electrical Engineers. Also published under the same arrangement are the Russian electronic journals RADIO ENGINEERING and TELECOMMUNICATIONS.

RADIO ENGINEERING AND ELECTRONIC PHYSICS

(РАДИОТЕХНИКА И ЭЛЕКТРОНИКА)

Publication of the Institute of Radio Engineering and Electronic Physics,
Academy of Sciences of the USSR

Translated and Produced
by
Royer and Roger, Inc.

Translation Editor: Leonard S. Schwartz

AIEE REVIEW COMMITTEE FOR RADIO ENGINEERING AND ELECTRONIC PHYSICS
Leonard S. Schwartz
New York University College of Engineering
Chairman

A.W. Bickley
W.P. Birkemeier
T.T.W. Bucher
J.L. Callahan
G.R. Cooper
W.A. Depp
R.S. Enticknap
A. Burr Fontaine
F.E. Froelich
Paul H. Gleichauf
G.S. Glinski
Bernard Harris
R.K. Hellmann
D.E. Higginbotham
B.A. Lengyel
W. Miller
Harry Rowe Mimno
W.W. Peterson
B. Reiffen
W.G. Schmidt
Herbert Sherman
D.L. Solomon
C.A. Stutt
G.Z. Sziklai
Joseph Vogelman
G.M. White
C.H. Wilcox
F.B. Wood
H.L. Yudkin

Subscriptions to Radio Engineering and Electronic Physics should be sent to AIEE
Special Subscription Department
41 East 28th Street, New York 16, New York

1961 Subscription rates:

	$	£
Individuals	28.50	10
Libraries, institutes, govt. agencies	57.00	20

12 issues per annum comprising approximately 1900 pages

EFFECTIVENESS OF SEARCH ALGORITHMS BASED ON THE USE OF TRIAL STEPS OF CONTROLLED DURATION

Yu. B. Kobzarev, A. Ye. Basharinov

Algorithms are considered for a search procedure in which the duration of trial steps is controlled. The indices characterizing the effectiveness of the controlled trial procedures are defined. Estimates are given of the influence of the resolving power (the manifold of signal forms) on the indices of controlled search procedures.

1. The search for the required (optimal) behavior (setting) plays an important role in automatic-measurement, servomechanisms, in self-adaptive systems, and also in automatic radio detection devices [1, 7, 8].

Simultaneous sampling of all possible behaviors (settings) leads as a rule to excessively cumbersome devices, and consequently the search process is subdivided into trial steps.

The effectiveness of a search procedure is estimated by means of indices that represent the operating reliability of the performance, as well as by a quantity that characterizes the "efforts" expended.

The efforts expended in search can be characterized by the time necessary to survey the specified range, the amount of energy consumed, etc.

Noise can cause the search operation to terminate in an untrue solution; the reliability of the operation may, for example, be estimated in terms of the values of the probabilities of the true and untrue solutions.

At the present time, little has been done towards synthesis of optimal search systems in the sense of minimizing the expended efforts at specified reliability indices, or in the sense of realizing the best indices for a fixed expended effort (see [1, 2, 8]).

It is obvious that a search method in which identical trial steps are used can in general deviate considerably from the optimal procedure, in view of the nonuniformity of the a priori distributions, the difference between the weights of the true and untrue solutions, etc.

Along with investigating general methods for the synthesis of optimal search systems, it is expedient to consider the feasibility of particular methods that increase the effectiveness of search procedures; such methods include the several variants, considered below, of the method of trial steps of controlled duration.

2. The duration of trial steps in a system admitting of a finite number of discrete modes can be controlled in various manners.

We shall consider three methods of controlling the duration of steps, connected with a sequential search procedure: a) use of statistical sequential analysis; b) use of grouped sequential procedures; c) selection of elements for subsequent repeated analysis.

In the first method, the data obtained in each trial step (element) are used to form a running likelihood coefficient. The values of the likelihood coefficient obtained in successive intervals of temporal quantization are compared with the thresholds, and when the upper or lower thresholds are reached the next trial step is made.

The second method differs in that the foregoing values of the likelihood coefficient and comparison with thresholds are carried out not at each quantization interval, but directly over the data of a given group of intervals. After comparing with the thresholds, a change is made to a new element or the termination of the given step is delayed by the time of the next group.

The third method of element selection is based on breaking up the given search time into stages. In the first stage all the elements are analyzed, and then some of the elements are sifted out; in the second stage the sampling is only with respect to the remaining elements, etc.

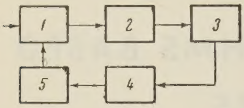


Fig. 1. Functional diagram of automatic tuning. 1 — mixer; 2 — accumulating device; 3 — comparison block; 4 — tuning-control unit; 5 — tunable heterodyne.

It is easy to see that the method of trial steps in which the duration is controlled with the aid of a sequential-analysis algorithm realizes the optimal procedure in a discrete system.

In fact, for a discrete system, the minimum expenditure of time (energy) necessitated by the survey of the aggregate of the elements is attained when the expenditures for the analysis of the individual elements is a minimum. However, according to the Wald-Wolfowitz theorem [5], for fixed reliability indices, the minimum duration for the element sounding procedure is attained when the data are processed by the method of likelihood coefficient with two-threshold selection.

Thus, a procedure employing trial steps of controlled duration, based on a sequential-analysis algorithm, is optimal in the aforementioned sense.

A feature of a search procedure with trial steps of controlled duration is the random character of the expenditure of efforts for the performance of the entire search operation.

An analysis of the method of trial steps of controlled duration has as its purpose a clarification of the dependence of the mean expenditure of efforts on the number of possible solutions (form of the signal), and also a determination of the statistical indices, connected with the random character of the expenditures (for example, the dispersion of the search time or of the energy expended).

3. The principles of construction of a search scheme with trial steps of controlled duration can be illustrated by using as an example a system for discrete automatic tuning of a receiver.

The functional diagram of such a system consists of a tunable selective element, a two-threshold analyzer, and an actuating device (see Fig. 1).

During the time of the trial step, the setting of the selective element is maintained constant until the comparison procedure is terminated; the actuating unit is then commanded to retune the selective element to a new position (or to change over from the search mode to the follow-up mode).

The following relations characterize the effectiveness, in the sense of reducing the search time of a procedure with trial steps of controlled duration.

The average time required to review n elements in the absence of a signal is

$$\bar{T}_{st} = n\overline{\Delta t}_{st}(D, F). \tag{1}$$

In the presence of a signal, in one of the elements, the time is

$$\bar{T}_{s\cdot st} = (n-1)\,\overline{\Delta t}_{st} + \overline{\Delta t}_{s\cdot st}(D, F), \tag{2}$$

where D is the required signal-detection probability and F is the permissible probability of false operation.

In the case of a search with a large number of elements ($\bar{T}_{s\cdot st} \simeq \bar{T}_{st}$), the effectiveness of the search will be characterized by the quantity

$$\eta_T = \frac{T_0}{\bar{T}_{st}} \simeq \frac{\Delta t_0}{\overline{\Delta t}_{st}}, \tag{3}$$

where Δt_0 is the duration of the trial step in the case of uniform sounding.

Owing to the reduced duration of the sounding, we can, as follows from [4], use the following estimate in the case of a two-threshold sequential procedure with $F \simeq 1-D$:

$$2 < \eta_T < 4,$$
$$\eta_T \to 4 \text{ for } F = (1-D) \to 0.$$

When $F \ll (1-D)$, as follows from [3],

$$\eta_T \simeq \frac{\ln F}{\ln (1-D)}. \tag{4}$$

The influence of randomness of the review time on the character of the process of the controlled search can be estimated from the magnitude of the standard deviation of the review time from its mean value.

By virtue of the independence of the duration of the procedure of sounding the individual elements, the relative magnitudes of the standard deviation of the survey time are determined from the relation

$$\delta_T = \frac{\sigma_T}{\overline{T}} = \frac{1}{\sqrt{n}} \frac{\sigma_{\Delta t}}{\overline{\Delta t}}, \tag{5}$$

where $\sigma_{\Delta t}$ is the standard deviation of the duration of sounding in the sequential procedure (see [3] and [6]; inasmuch as $\sigma_{\Delta t} < 3\overline{\Delta t}$ in the ranges of practical interest, we have $\sigma_T < 0.3\overline{T}$ when $n > 100$.

When trial steps of controlled duration are used, the search time can be reduced through economy in power expenditure (for example, through reduction in threshold-signal intensity).

The redistribution of the duration of the trial steps over the signal on and off intervals makes it possible to increase the duration of the sounding of an element containing a signal, while retaining the total search time constant.

Let us give the relations for the average duration of the trial steps for a constant total search time.

Solving the equation with respect to $\overline{\Delta t}^*_{\text{s. st}}$ we get

$$\begin{gathered}(n-1)\,\overline{\Delta t}^*_{\text{st}} + \overline{\Delta t}^*_{\text{s. st}} = n\Delta t_0, \\ \overline{\Delta t}^*_{\text{s. st}} = \xi\,\overline{\Delta t}^*_{\text{st}}\end{gathered} \tag{6}$$

[$\xi = f(D, F)$ is the coefficient of asymmetry of the sequential procedure], we obtain

$$\frac{\overline{\Delta t}^*_{\text{s. st}}}{\Delta t_0} = \frac{\xi n}{(n-1)+\xi}; \tag{7}$$

we denote

$$\zeta = \frac{\Delta t_0}{\overline{\Delta t}_{\text{s. st}}}; \tag{8}$$

$\overline{\Delta t}_{\text{s. st}}$ is the average duration of the trial step in the presence of a signal with threshold intensity corresponding to the intensity for uniform trial steps. The quantity $\zeta = \Delta t_0/\overline{\Delta t}_{\text{s. st}}$ is greater than 1[$\zeta \simeq 1/D$ when $F \ll (1-D)$]. Using (7) and (8), we obtain a relation characterizing the gain in time required to sound an element containing a signal.

$$\begin{gathered}\eta_t = \frac{\overline{\Delta t}^*_{\text{s. st}}}{\overline{\Delta t}_{\text{s. st}}} = \zeta \frac{\xi n}{(n-1)+\xi}, \\ \eta_t \simeq \zeta\xi \text{ for } n \gg 1.\end{gathered} \tag{9}$$

Let us estimate the reduction in the intensity of the threshold signals in the case of controlled search; this reduction is the result of redistribution of the average durations of the trial steps in the case of coherent and incoherent reception.

In the case of coherent accumulation, i.e., in the case of optimum utilization of the signal energy, it follows from the relations cited in [6] that

$$q^2_{\text{coh}} = \frac{2\left[D \ln \frac{D}{F} + (1-D) \ln \frac{1-D}{1-F}\right]}{\overline{\Delta t}_{\text{s. st}}} \tag{10}$$

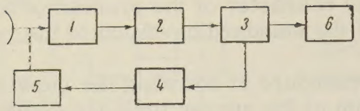


Fig. 2. Functional diagram of apparatus with controlled sounding mode.

1 — Transmitting and receiving devices; 2 — multichannel accumulator; 3 — logical decision element; 4 — command block; 5 — beam position control element; 6 — marker.

where $q^2_{\text{coh}} = \mathcal{E}/N_0$ is the ratio of the energy obtained in single sampling to the spectral density of the noise N_0. From (9) and (10) we obtain

$$\eta_Q = \frac{q^2_{\text{coh}}}{(q_{\text{coh}})^2} = \eta_t. \tag{11}$$

In incoherent accumulation of weak signals we have by virtue of the losses due to suppression of the signal, as follows from [6],

$$q^4_{\text{inc}} \simeq \frac{2\left[D \ln \frac{D}{F} + (1-D) \ln \frac{1-D}{1-F}\right]}{\overline{\Delta t}_{\text{s. st}}}, \tag{12}$$

$$\eta_Q \simeq \sqrt{\eta_t}. \tag{13}$$

Recognizing that when $F \ll D$ we have

$$\xi = \frac{D \ln \frac{D}{F} + (1-D) \ln \frac{1-D}{1-F}}{F \ln \frac{D}{F} + (1-F) \ln \frac{1-D}{1-F}} \simeq \frac{D \ln \frac{D}{F}}{\ln (1-D)}, \tag{14}$$

we obtain estimates for the energy efficiency of the search procedure in the case of asymmetrical equivalent values of erroneous solutions, in the form

$$\sqrt{\frac{\ln F}{\ln (1-D)}} \leqslant \eta_Q \leqslant \frac{\ln F}{\ln (1-D)}. \tag{15}$$

The asymptotic values of the upper and lower estimates in the symmetrical case, when $(1-D) = F \to 0$, have according to (3), (11), and (13) the following form

$$2 \leqslant \eta_Q < 4.$$

4. An important example of the application of the method of trial steps of controlled duration is the radar survey scheme [8].

The antenna beam executes intermittent soundings, in which the exposure in each direction is controlled by a two-threshold system (Fig. 2).

After deciding to shift the beam to a new direction, the accumulating device is cleared and the process is repeated with the new beam position.

A feature of the controllable search mode is in this case the presence of resolution in the coordinates and the parameters in each trial step; this is equivalent to a variety of forms of received echo signal, and is connected with the fact that the receiver contains many channels.

Let us examine the relations characterizing the efficiency of the controlled survey in this case.

Let the total survey zone consist of elementary sectors. The average time of survey of the zone in the absence of a signal is

$$\overline{T}_{\text{st}} = n\overline{\Delta t}_{\text{st}}(m).$$

In the presence of s reflectors, the average survey time is

$$\overline{T}_{s.\,st}(s) = (n - s)\,\overline{\Delta t}_{st} + \sum_{i=1}^{6} \overline{\Delta t_i} \tag{16}$$

When $s \ll n$ we have $\overline{T}_{s.\,st}(s) \simeq \overline{T}_{st}$, so that in the case of a sufficiently large number of elements in the survey zone the total average survey time is determined by the time required to survey the "empty" directions.

The economy in sounding time, compared with the method of uniform uncontrolled survey with equivalent reliability indices, is determined by the relations (9).

Let us rewrite (9) for a m-channel system in the form

$$\eta_t(m) = \eta_t(1)\frac{1}{\varphi(m)}, \tag{17}$$

where

$$\varphi(m) = \frac{\psi(m)}{\varkappa(m)}; \quad \psi(m) = \frac{\overline{\Delta t}_{st}(m)}{\overline{\Delta t}_{st}(1)}; \quad \varkappa(m) = \frac{\Delta t_0(m)}{\Delta t_0(1)}$$

The average duration of the sequential procedure in a multichannel case has, according to [9], a linearly logarithmic dependence on the number of channels m:*

$$\overline{\Delta t}_{st}(m) = \overline{\Delta t}_{st}(1)\,[1 + k \ln m]. \tag{18}$$

The dependence of the intensity of the threshold signals on the number of channels was considered, for the single-threshold procedure, in several papers (see, for example, [10]). According to [10]

$$q^2_{inc}(m) = q^2_{inc}(1)\,(1 + \gamma \ln m), \tag{19}$$

where γ depends on the form of the signal (for example, in the case of a Rayleigh signal we have $\gamma \simeq \frac{1}{\ln\frac{1}{F} - \ln\frac{1}{D}}$).

In coherent accumulation, the coefficient $\varkappa(m) = q^2_{coh}(m)/q^2_{coh}(1)$; in incoherent accumulation we have $\varkappa(m) \simeq [q^2_{inc}(m)/q^2/_{inc}(1)]^x$, where $1 < x \leq 2$. The energy efficiency of the method of trial steps of controlled duration can be estimated with the aid of (12) and (13).**

By way of example, Table 1 lists the calculated values of estimates of the coefficients η_Q for a false-alarm probability $F = 10^{-6}$.

Table 1

m	Coherent Mode		Incoherent Mode	
	D=0.5	D=0.9	D=0.5	D=0.9
1	19.8	6	4.5	2.46
10	7.2	3	2.95	1.9
100	4.95	2.2	2.67	1.76
1000	4.05	1.86	2.56	1.72

5. To simplify the search procedure, we can use a method wherein the duration of the trial step is controlled by means of data obtained at the ends of fixed intervals (grouping intervals). The number of independent data in the grouping interval is obviously equal to the ratio $\Delta_{gr}/\Delta\tau$, where $\Delta\tau$ is the interval of the temporal quantization.

---

*The method of statistical experiment (see [9]) yields for the coefficient k the estimates $k \simeq 1$ when $D \leq 0.5$ and $0.5 < k < 1$ when $D > 0.5$.

**Estimates of the efficiency of controlled survey by means of relations (9), (17), and (18) are in sufficiently good agreement with the data given in [8].

The use of the grouping method leads to a discrete variation in the duration of the trial step by an amount which is a multiple of the grouping interval.

In the grouped procedure method, the average duration of the trial steps can be shown not to exceed the average duration of the trial step attained in sequential analysis of the data, in the worst case, by more than twice the grouping interval [5].

The following estimate can be employed with a high degree of probability

$$\overline{\Delta t}_{\text{gr}} \leqslant \overline{\Delta t} + \Delta_{\text{gr}}. \tag{20}$$

Using (20), we obtain the condition for the choice of the grouping interval

$$\Delta_{\text{gr}} < (0{,}2 - 0{,}3)\, \overline{\Delta t_{\text{st}}} \tag{21}$$

If condition (21) is satisfied, we can neglect the "losses" connected with the use of the grouped procedure.

When the duration of the trial steps is controlled by means of the grouped data, the search procedure can be represented in the form of an equivalent multistage process: a) at the first stage all the elements are sounded with a trial-step duration equal to the grouping interval, and two-threshold comparison is used to select the elements for the repeated sounding; b) at the second and succeeding stages, the sounding process with duration Δ_{gr} is repeated on those elements on which no solution was obtained in the preceding stage; c) single-threshold comparison is carried out in the concluding stage.

Using grouping intervals equal to the temporal quantization intervals, we obtain a multistage search process which is equivalent to a process with continuous control of the trial-step duration.

This premise can be formulated as a statement that multistage search processes are energetically equivalent to a search with continuous control of the duration of the trial steps.

6. One of the variants of multistage construction of a search procedure is a method based on single-threshold selection.

The algorithm for the multistage procedure can be represented in the following fashion: a) at the first stage the results of the uniform sounding are used to pick out elements in which the output voltage exceeds the threshold A_1; b) only the selected elements (i.e., elements in which $u_i > A_1$) are sounded in the second stage, and from these elements are picked in which the voltage exceeds the threshold A_2, etc. The number of sounding stages, k, is fixed beforehand.

The reliability of the search procedure is characterized by the following indices:
D — probability of correct detection of the signal

$$D = \prod_{i=1}^{k} D_i, \tag{22}$$

F — probability of false alarm

$$F = \prod_{i=1}^{k} F_i. \tag{23}$$

By virtue of the random nature of the number of elements selected in each stage, the energy expenditure for each individual operation is a random quantity

$$Q_k = \sum_{i=1}^{k} l_i Q_i (D_i, F_i), \tag{24}$$

where l_i is the number of elements sounded at the i-th state (the number l_i has a binomial distribution); $Q_i(D_i, F_i)$ is the threshold ratio of the signal energy to the spectral density of the noise.

The average energy expenditures in the absence of a signal are determined by the relation

$$\overline{Q}_k = l\left[Q_1 + F_1Q_2 + F_1F_2Q_3 + \ldots + \prod_{i=1}^{k-1} F_i\, Q_k\right]. \tag{25}$$

For a fixed number of stages, the choice of threshold values and the distribution of energies over the stages can be determined from the condition that the average energy expenditures necessary to ensure some specified reliability indices be a minimum.

The reduction in the energy expenditure, compared with the case of uniform single-stage survey, can be estimated in terms of the coefficient $\eta_Q = Q_1(D, F)/\overline{Q}_k(D, F)$, where $Q_1(D, F)$ is the energy expended in uniform search.

Let us consider examples of selection of an optimal distribution of expenditures in the case of a two-stage search procedure. The problem of choosing the optimal distribution reduces in this case to finding a conditional extremum of (25) with respect to the variables D_1 and F_1.

a) Search of a signal of fully known form, expected in one among l possible positions. The expected signal is a sinusoidal oscillation with known amplitude.

The energy expenditure, expressed in terms of the threshold ratio of the signal energy to the spectral density of the noise, can according to (25) be represented in the form

$$\overline{Q}_2\,(D, F) = l\,[(\lambda_{D_1} + \lambda_{F_1})^2 + F_1\,(\lambda_{D_2} + \lambda_{F_2})^2], \tag{26}$$

where λ_α is the quantile of the normally distributed quantity

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{\lambda_\alpha}^{\infty} e^{-\frac{x^2}{2}}\, dx, \tag{27}$$

$$D = D_1D_2; \quad F = F_1F_2.$$

The optimum values of the parameters D_1 and F_1 are determined from the equations

$$\frac{\partial \overline{Q}_2}{\partial D_1} = 0, \quad \frac{\partial \overline{Q}_3}{\partial F_1} = 0. \tag{28}$$

Eq. (28) was solved graphically with account of (27). Table 2 lists the values of the coefficient η_Q, estimating the gain above the single-stage uniform search with optimal choice of the parameters D_1 and F_1.

Table 2

D_1	F_1		
	10^{-4}	10^{-6}	10^{-8}
0.5	3.38	4.74	6.36
0.9	2.46	3.58	4.13

b) Search of fluctuating signals with Rayleigh distribution of intensity, expected in one among l possible positions, with m-channel reception in each position. The relative intensity of the threshold signals can in this case, as shown in [10], be represented with good accuracy, under optimal processing, by the following relation

$$Q_2 = \frac{|\ln F| - \ln m}{|\ln D|} - 1. \tag{29}$$

Eq. (26) for the energy expenditure assumes, with account of (27) and (29), the following form

$$\overline{Q}_2 = l\left[\left(\frac{|\ln F_1| - \ln m}{|\ln D_1|} - 1\right) + F_1\left(\frac{\left|\ln \frac{F}{F_1}\right| - \ln m}{\left|\ln \frac{D}{D_1}\right|} - 1\right)\right]. \tag{30}$$

The equations for the choice of optimal values of D_1 and F_1 can be obtained by using (28) and (30).

The numerical values of the coefficient η_Q, corresponding to optimal values of D_1 and F_1, are listed in Table 3.

Table 3

D_1	F_1					
	10^{-4}	10^{-6}	10^{-8}			
	$m=1$	$m=1$	$m=1$	$m=10$	$m=100$	$m=1000$
0.5	2.05	2.54	3.2	2.45	2.0	1.78
0.9	1.78	2.37	2.96	2.36	1.95	1.75

A comparison of the data listed in Tables 2 and 3 shows that the two-stage search mode is more efficient (compared with the uniform mode) in the case when signals of regular intensity are observed, rather than in the case of fluctuating signals.

An increase in the resolving power, as in the case of two-threshold control of the duration of the trial steps (see Table 1), smooths out the differences in the indices of the uniform and controlled search modes.

CONCLUSION

We have shown that controlled search can ensure a reduction in the expenditures entailed in signal detection. In particular, controlled directional search makes it possible to reduce the average radiated power for a specified total survey time.

The manifold possible signal forms (brought about by the resolution in delay time and by the Doppler shift, or by fluctuations in the signal intensity) reduces the gain due to the application of controlled search.

REFERENCES

1. A.A. Fel'dbaum. Computers in Automatic Systems, GIFML, 1959.
2. B. Koopman. The theory of search, Operations Research, 1957, 5, 5.
3. A.E. Basharinov, B.S. Fleyshman. On the efficiency of the method of sequential analysis in devices for the detection of weak signals in noise, Radiotekhnika i elektronika, 1958, 3, 6, 835.
4. S.A. Ayvazyan. Comparison of optimal properties of the Neyman-Pearson and Wald criteria. Theory of Probability and Its Applications. 1959, 4, 1, 86.
5. A. Wald. Sequential Analysis, N.Y., Wiley. 1947.
6. D. Middleton, J. Bussgang. Optimum sequential detection of signals in noise, IRE Trans., 1955, IT-1, 3, 5.
7. H. Blasbalg. The sequential detection of a sine-wave carrier of arbitrary duty ratio in Gaussian noise, IRE Trans., 1957, IT-3, 4, 248.
8. G. Preston. The search efficiency of sequential probability ratio search radar, Internat. Convention Record, 1960, p. 4, 116.
9. A.E. Basharinov, B.S. Fleyshman, G.S. Tyslyatskiy. Method of sequential analysis in problems involving the detection of signals in multi-channel systems. Paper delivered at the Jubilee Session of the A.S. Popov All-Union Scientific Technical Society for Radio Engineering and Electric Communication, 1959. Collection for the A.S. Popov Centennial Celebration, Academy of Sciences Press, 1960, pp. 74-78.
10. L.A. Vaynshteyn, V.D. Zubakov. Separation of Signals from a Background of Random Noise. Soviet Radio Press, 1960, p. 359.

Received by editor
7 October 1960

RADIO ASTRONOMIC OBSERVATIONS OF SOVIET SPACE ROCKETS

V.V. Vitkevich, A.D. Kuz'min, L.I. Matveyenko,
R.L. Sorochenko, and V.A. Udal'tsev

We describe the use of a radio interference method for the reception of radio signals from the first three Soviet space rockets. The fundamental equations of the radio interference method of observation are derived. The apparatus and the procedure used for aligning the radio interferometer against discrete radio emission sources are described. The time and the coordinates of the landing of a container with scientific apparatus from the second Soviet space rocket on the moon are determined.

INTRODUCTION

One of the problems involved in the study of outer space with the aid of rockets is the determination of the angular coordinates of a space rocket, i.e., its position on the celestial sphere. Usually the space rocket (in the container with scientific apparatus) carries a radio transmitter, so that these measurements can be carried out by radio methods. The problem of determining the coordinates of a radiating source in outer space has an analogy in radio astronomy, where the angular coordinates of discrete radio emission sources are determined by using for this purpose the radio-interference method [1], which has good accuracy.

We have used the same method to observe the Soviet space rockets, and were thus also able to determine, with a good degree of accuracy, the intensity flux of radio emission from transmitters mounted on space rockets.

1. FUNDAMENTAL EQUATIONS OF THE RADIO INTERFERENCE METHOD OF OBSERVATION

The ordinary radio interferometer used in radio astronomy consists of two separated antennas, the signals from which are added, amplified, and recorded. If β is the angle between the direction to the signal source and the normal to the interferometer base, ω is the angular frequency of the incoming signal, D the interferometer base (distance between antennas), and l_1, l_2 the electric lengths of the cables from each antenna to the place where the signals are added, then the voltage at the receiver input, for identical antennas, is determined, apart from a constant factor, by the following expression

$$u_{\mathrm{in}} = \cos(\omega t + \psi - kl_1) + \cos(\omega t + \psi - kl_2 - kD\sin\beta), \tag{1}$$

where ψ is the initial phase of the incident wave, and $k = 2\pi/\lambda$; and λ is the wavelength in free space.

The voltage at the receiver output, in the case of square-law detection, is proportional to

$$u_{\mathrm{out}} = \left\{1 + \cos\left[\frac{2\pi D\sin\beta}{\lambda} + \frac{2\pi(l_2 - \lambda_1)}{\lambda}\right]\right\}. \tag{2}$$

The angular width of one lobe of the interference pattern, $\Delta\beta$, defined as the distance between two neighboring zeroes, is given for the case of small β by the relation

$$\Delta\beta = \frac{\lambda}{D\cos\beta}. \tag{3}$$

When $\beta \simeq 0$ and $D = 100\,\lambda$ we have, for example $\Delta\beta = 34.4'$.

The expression (2) contains the difference in length of the electric cables, $l_1 - l_2$. This difference determines the initial phase of the interference pattern. If the antennas have different phase characteristics, the interference pattern can also shift, a fact accounted for by introducing corresponding corrections to l_1 and l_2. Having the interference pattern, i.e., the dependence of u_{out} on the time, and knowing the values of D, λ, and $l_2 - l_1$, we can determine the angle from expression (2). This angle, however, is not determined uniquely and is accurate only to within the width of one lobe of the interference pattern. This angle is easiest to determine at instants corresponding to the maxima of the interference pattern. The quantity β_{max} is determined in terms of the number n of the lobe and the interferometer parameters by the following relation:

$$\sin \beta_{\mathbf{max}} = \frac{\lambda}{D}(n - \eta). \tag{4}$$

where η is a parameter that depends on the difference in electric distances from each antenna to the place where the signals are added [see Eq. (11), below]. When the electric lengths are equal we have $\eta = 0$.

The azimuth of the source A is connected with the angle β by the relation

$$\sin\beta = \sin\gamma\cos z + \cos\gamma\sin z\sin(A - \theta), \tag{5}$$

where z is the zenith angle of the source, γ the angle between the horizontal plane and the projection of the base on the vertical plane passing through the east-west line, and θ the angle between the east-west line and the projection of the base on the horizontal plane. Then

$$A = \theta + \arcsin\left[\frac{1}{\cos\gamma\sin z}\frac{\lambda}{D}(n - \eta) - \operatorname{tg}\gamma\operatorname{ctg} z\right]. \tag{6}$$

2. APPARATUS

Basic principles. We have used a specially developed radio interferometer with phase modulation. To separate the signal-amplitude variations due to changes in the position of the source relative to the interference-pattern lobe from the change in the intensity of the source itself, we used the principle of phase-shifted two-channel reception [2]. When the signal received has a continuous spectrum, the fluctuation sensitivity, in antenna temperature units (T_a), is given by the well known relation *

$$\delta T_{\mathrm{a}} = a_1 T_0 F_{\mathbf{eq}}\sqrt{\frac{1}{\Delta f\tau}}, \tag{7}$$

where a_1 is a dimensionless factor that depends on the characteristics of the receiver (in general, with error not greater than a factor of 2, we have $a_1 = 1$); T_0 is the standard temperature of the ambient; $F_{eq} = (T_a + T_i)/T_0$ is the equivalent noise coefficient; T_i is the reduced noise temperature at the receiver input, due to the intrinsic noise, $T_i = T_0(F_{rec} - 1)$; F_{rec} — noise factor of the receiver; T_a — antenna temperature**; τ — accumulation time of output apparatus; Δf — bandwidth of the receiver up to the detector.

If the signal received is monochromatic or sufficiently narrowband (compared with the bandwidth of the receiver), it can be shown that the fluctuation sensitivity is determined

---

*Expression (7) is taken from [3]. For the case of interest to us we have assumed $T = T_a$, $\Delta\Omega \gg \Delta\Omega_1$, $\overline{\alpha^2} = 0$ and $\Delta\Omega_1 = \pi/2\tau$.

**In the meter band, the receiver input noise can be due to a large extent to galactic radiation [4]. Thus, at 200 Mc the brightness temperature in the direction towards the center of the galaxy is about 1500°K. This means that the noise temperature in the case of observations near the center of the galaxy can be increased by this amount, with corresponding reduction in sensitivity.

(in power units at the receiver input) by the following relation

$$\delta P_{\text{in}} = a_2 k T_0 F_{\text{eq}} \sqrt{\frac{\Delta f}{\tau}}, \tag{8}$$

where k is Boltzmann's constant and a_2 is a coefficient close to unity. It is seen from the foregoing equations that to improve the fluctuation sensitivity to a signal with continuous spectrum it is more convenient to operate with a greater bandwidth, whereas a narrow band is more suitable for monochromatic signal. In our observations we alternately received both types of signals. For this reason, both requirements were taken into account in choosing the parameters for the receiver.

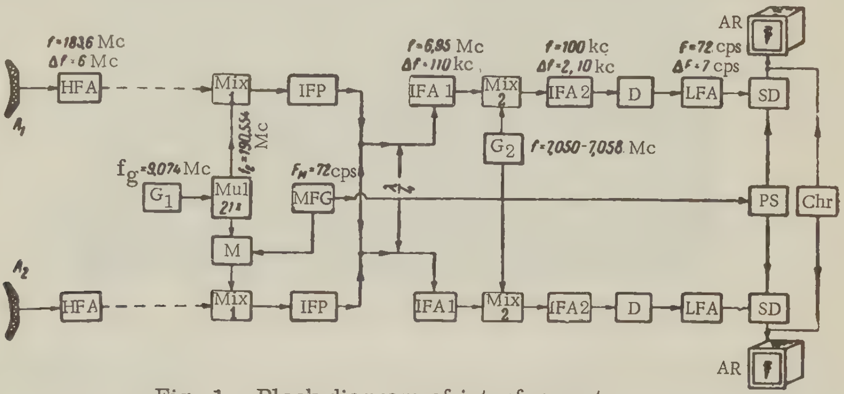


Fig. 1. Block diagram of interferometer

HFA—High frequency amplifier; Mix—Mixer; IFP—IF Preamplifier; IFA—IF Amplifier; D—Detector; LFA—Low frequency amplifier; AR—Automatic recorder; SD—Synchronous detector; G—Generator; Mul—Multiplier; MFG—Modulating frequency generator; PS—Phase shifter; Chr—Chronometer.

Block diagram of radio interferometer. The observations were carried out at 183.6 Mc, corresponding to the frequency of the transmitter mounted in the capsule with scientific apparatus. A block diagram of the radio interferometer is shown in Fig. 1. The interferometer antennas were two truncated parabolic reflectors measuring 8 x 18 and 11 x 22 meters (Fig. 2), spaced 176 meters apart in the east-west direction. The focal distance of the two antennas was 8 meters. One of the antennas was controlled by a program device which automatically tracked the sources with constant direct ascension α and declination δ. The second antenna moved in synchronism with the first. This mode was used to align the radio interferometer against the discrete sources. In observations of space rockets, the antenna tracking was semi-automatic in accordance with a prescribed program.

Specially developed radiating systems* were located in the foci of the paraboloids and ensured optimum irradiation of the mirrors in two mutually perpendicular linear polarizations. The radiating system for the horizontal polarization was a three-stack in-phase antenna with one element per stack and with untuned counter reflector. The radiating system for the vertical polarization was in the form of an inphase two-stack antenna with two elements per stack. A sketch of the radiating system is shown in Fig. 3. The main element of the system is the half-wave dipole. The distance between dipoles was determined by calculation and trimmed using a breadboard, for the 10 cm wavelength. The breadboard was also used to measure the mutual decoupling of the radiating systems of the different polarizations; this decoupling was not less than 30 db.

The three dipoles of the horizontal-polarization radiator were connected by means of equal segments of RK-20 cable to the common supply feeder. The two outside dipoles were made symmetrical with the aid of a quarter-wave short-circuited stub of two-conductor line, which simultaneously served as a quarter-wave mounting insulator. The central dipole, which was folded, was made symmetrical with the aid of a U-bend. Matching with the common line was with the aid of a loop. The traveling wave coefficient of the radiating system was 0.75 -- 0.8.

---

*Yu. P. Ilyasov participated in the development of the radiating system.

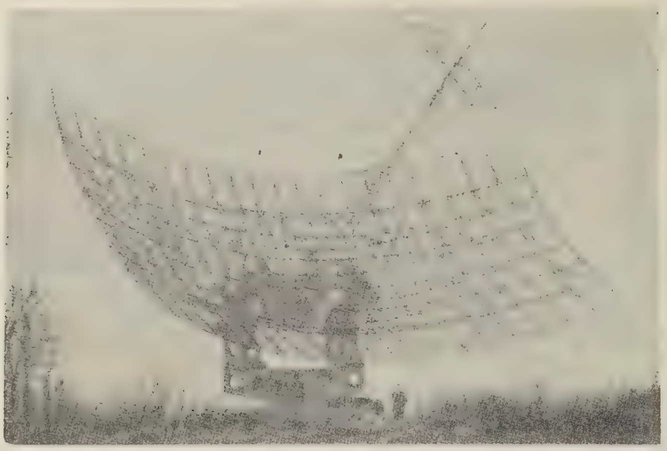


Fig. 2. Over-all view of one of the interferometer antennas

The directivity patterns and utilization coefficients of the antenna surfaces were measured against the radio emission from the sun. The surface utilization coefficient was about 0.5 for each antenna.

Phase modulation at a frequency of 72 cps was effected by reversing the phase of the signal through periodic variation of the electric length of the cable between the heterodyne and one of the mixers. This caused "rocking" of the interference pattern of the radio reception, and the received signal was thus amplitude modulated at this frequency. This modulation system ensures good interference immunity.

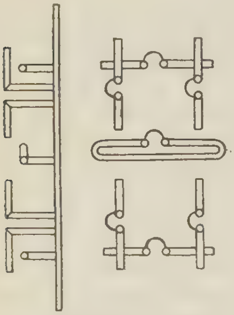


Fig. 3. Sketch of radiating system.

The phase modulator employed the customary hybrid switch circuit. The switching elements used were beam diodes 6D3D, controlled by sinusoidal modulating voltage. The use of these diodes kept the power attenuation of the heterodyne below 2 db. By adjusting the slope of the diode characteristic by changing the filament voltage, and also by connecting the modulator in the heterodyne circuit, it was possible to reduce greatly the parasitic amplitude modulation, compared with previously employed schemes [5]. When the modulated voltage amplitude is sufficiently large, the phase modulation of the heterodyne voltage is close to rectangular. This modulator permits converting the phase-modulated radiometer into an amplitude-modulated radiometer, by turning off the control voltage applied to one of the diodes. Amplitude modulation was used in the alignment of the individual antennas of the interferometer.

Preamplification of the signal was in the UHF blocks mounted directly on the antennas. The noise factor of the UHF blocks was 5. Amplified signals from each antenna were converted, passed through buffer stages, and summed at the first IF (6.95 Mc), after which they are split among two channels shifted 90° apart.

As indicated above [see Eq. (8)], for a monochromatic signal one obtains better fluctuation sensitivity if a narrow bandwidth is used. In this connection, double frequency conversion and crystal stabilization of the two local heterodynes is employed in the apparatus. The first heterodyne, whose frequency (190.554 Mc) is close to the received-signal frequency, is most important for the over-all frequency stability of the setting of the apparatus. Its frequency is set by a thermostatically controlled crystal heterodyne, whose frequency (9.074 Mc) is subsequently multiplied 21 times. The first heterodyne has a relative instability of

10^{-6}, which dictated the choice of a two kc bandwidth for operation with a monochromatic signal.

To ensure reception when the frequency of the received signal is changed by the Doppler effect, provision is made in the apparatus for a fixed stepwise retuning of the receiver frequency within limits of 8 kc. For this purpose, the second heterodyne contains five crystal controlled master resonators, spaced 2 kc apart in frequency. In addition, the first heterodyne can be continuously tuned within the limits ± 3.2 kc.

When a continuous-spectrum signal is received, as during the calibration and alignment of the apparatus against the discrete radio-emission sources, it is desirable to have a greater bandwidth [7]. For this purpose, provision is made in the receiver for the possibility of fixed switching of the band of the second IF amplifier from 2 to 10 kc, retaining the same tuned frequency and gain. The signal separated by the synchronous detector is accumulated on the output RC filter. The accumulation time τ is limited by the period of the interference record. If τ is commensurate with the period of the interference, the interference record becomes distorted (the amplitude is decreased and the maxima and minima are shifted). In our case the accumulation time was chosen to be 26 sec, so as to ensure the necessary fluctuation sensitivity without causing any practical changes in the amplitude of the interference record.

Our measurements have shown that the chosen accumulation value also ensures registration of periodic measurements of signals from the space rocket as the orientation of the latter is varied. To ensure stable operation of the apparatus (to ensure a constant gain and constant phase characteristics) and to attain high sensitivity, the filament and plate supplies of all the operating tubes of the receiver were fed from stabilized sources with a stabilization coefficient of the order of 1,000. The signal was registered with electronic type EPP-09 automatic recorders. An exact time scale in the form of minute markers was produced on the chart simultaneously with the signal.

Estimates of experimental sensitivity and accuracy of angle measurements. The experimental sensitivity of the receiver is listed in Table 1 as a function of the bandwidth. The experimentally obtained sensitivity is approximately half that calculated by formulas (7) and (8), this being due to various parasitic influences in the apparatus, which are not analyzed in the present paper.

The sensitivity of the interferometer with respect to the energy flux density is determined by the relations

$$p_{\mathrm{m}} = \frac{2k\delta T}{A_{\mathrm{ef}}}, \quad p = \frac{2\delta P_{\mathrm{in}}}{A_{\mathrm{ef}}\Delta f},$$

where $A_{\mathrm{ef}} = 190\ \mathrm{m}^2$ is the effective area of the antennas*; p_{m} is the energy flux density of the monochromatic signal; p is the energy flux density of the continuous spectrum (per unit frequency interval).

The values of the minimum observed energy fluxes are indicated in Table 2.

It is very important to determine the apparatus phase errors of various elements of the interferometer, since these errors determine the accuracy of the angle measurements. After differentiating Eq. (6) and after several transformations we obtain a general expression for the over-all error

$$dA = dA_1 + dA_2 + dA_3 + dA_4, \tag{9}$$

where

$$dA_1 = \left(\operatorname{tg} A + \frac{1}{\sin z}\,\frac{l_2 - l_1}{D}\right)\frac{d\lambda}{\lambda} \tag{9a}$$

---

*In the case of two different interferometer antennas and in the case of square-law detection, both antennas are equivalent to a single antenna whose area is equal to twice the average geometric area of the individual antennas, $E_{\mathrm{eq}} = 2\sqrt{A_1 A_2}$. In our case, where the surface utilization coefficient is 0.5 and the geometric areas of the antennas are 144 and 240 square meters, the effective area is 190 square meters.

Table 1

Signal	Bandwidth, kc	
	2	10
With continuous spectrum, T, deg.	16	7
Monochromatic, δ P, watt	$4.5\cdot10^{-19}$	10^{-18}

Table 2

Signal	Bandwidth, kc	
	2	10
Monochromatic, p_m, w/m^2,	$4\cdot10^{-21}$	10^{-20}
With continuous spectrum, p, w/m^2-cps	$2\cdot10^{-24}$	10^{-24}

is the error due to the frequency instability of the received signal when the cables to the first and second antennas are of unequal electrical length:

$$dA_2 = \frac{1}{\sin z \cos A}\,\frac{d(l_2 - l_1)}{D} \qquad (9b)$$

is the error due to instability of the electric lengths of the cables;

$$dA_3 = \frac{1}{\sin z \cos A}\,\frac{\lambda}{D}\,\frac{d(\varphi_1 - \varphi_2)}{2\pi} \qquad (9c)$$

is the error due to instability of the phase characteristics of the apparatus; and

$$dA_4 = \frac{1}{\sin z \cos A}\,\frac{\lambda}{2D}\,\frac{d\Delta}{2\pi}, \qquad (9d)$$

is the error due to instability of the phase modulator. Quantitative estimates of these errors show that in the presence of a crystal-controlled heterodyne, a high frequency buried in a trench, broadband high-frequency channel up to the point where the signals are added, and a stable modulator, the summary error does not exceed 1 or 2 minutes of angle.

The sensitivity obtained enables us to register a signal from a space rocket as far away as the moon, using a transmiter power of 0.03 w, and to make angle measurements accurate to 1-2'. Such apparatus parameters have been confirmed by reception of radio emission from the discrete source in the Hydra constellation (Fig. 4) the intensity of which is $2.5\cdot10^{-24}w^{-2}$ cps^{-1} [6], and a large cycle of alignment records obtained during the working process.

Fig. 4. Sample of recording of radio emission from the discrete source A-Hydrae

3. ALIGNMENT OF THE INTERFEROMETER

In carrying out the alignment,* a distinction must be made between the parameters that remain constant and those that vary in time. Among the constant parameters are γ, θ, and D, which are determined by the fixed geometry of antenna location. These parameters need be determined only once and can be used in all further observations. The parameters $\eta = \varphi_n/2\pi$, determined by the electric lengths of the cables and by the phase characteristics of the input stages of the receiver, may vary in time and should therefore be determined before each observation.

The values of D, γ, and θ were determined by a geodetic survey which yielded the following results: D = 175.896 m; $\gamma = 2°44'$; $\theta = -14'$. Substituting these values of D, γ, and θ, and $\lambda = c/183.6 \times 10^6$ into (6) we obtain

$$A = 179°46' + \arcsin\left[\frac{0{,}0093006}{\sin z}(n - \eta) - 0{,}047669 \operatorname{ctg} z\right]. \qquad (10)$$

To determine the parameter η it is necessary to align the directivity diagrams of the radio interferometer against a radio-emitting source located in the far zone of the interference antenna system, moving in space with an angular velocity close to the angular velocity of the space rocket; the coordinate of the source at each given instant can be determined with high degree of accuracy. We used as a source the discrete sources of radio emission A-Cassiopeiae, A-Cygni, A-Tauri, and A-Virginis, for which the equatorial coordinates (epoch 1959.5) and radio emission fluxes were assumed as listed in Table 3.

---

*V.M. Gorelova participated in the development of the alignment procedure.

Table 3

Radioemission Source	$\alpha_{(1959.5)}$	$\delta_{(1959.5)}$	P, $w/m^2 \cdot$ cps
A-Cassiopeiae	$23^h21^m38^s$	$+58°35'$	$120 \cdot 10^{-24}$
A-Cygni	$19^h58^m05^s$	$+40°36',5$	$70 \cdot 10^{-24}$
A-Tauri	$5^h32^m09^s$	$+21°59'$	$18 \cdot 10^{-24}$
A-Virginis	$12^h28^m47^s$	$+12°34'$	$9 \cdot 10^{-24}$

The parameter η was defined as the difference between the calculated time of passage of the source through the maximum of the interference diagram (t_{calc}) and the true time $t_{\Lambda\,true}$, referred to the period T of the interference lobe:

$$\eta = \frac{t_n - t_{\Lambda\,\mathrm{true}}}{T}. \tag{11}$$

Owing to the finite time constant of the output apparatus, the time of passage of the source through the maximum of the interference pattern, $t_{\Lambda\,true}$, does not coincide with the time t_Λ corresponding to the maximum declination of the registering instrument. The problem of determining the displacement $t_\Lambda - t_{\Lambda\,true}$ can be reduced to a determination of the phase shift of a sinusoidal voltage of frequency $\Omega = 2\pi D \sin \beta/\lambda$, passing through an integrating network with a time constant $\tau = RC$. It is easily shown that the phase shift is in this case

$$\psi = \mathrm{arctg}\,(\Omega RC).$$

Expanding $\tan^{-1}(\Omega RC)$ in a series and retaining the first two terms, we obtain

$$\Delta\tau = t_\Lambda - t_{\Lambda\,\mathrm{true}} = \tau\left[1 - \frac{4\pi^2}{3}\left(\frac{\tau}{T}\right)^2\right]. \tag{12}$$

This time correction must be introduced even when the time constant of the source observation is the same as the alignment time constant, for in general the observation and alignment are carried out at different elevation angles of the radiation source. Consequently, the interference period T obtained during alignment is different from that obtained during the observations, and this in turn brings about a different time shift $\Delta\tau$. A sample of the alignment record made on 13 September 1959 with the aid of the cosmic source A-Cygni is shown in Fig. 5.

4. MEASUREMENT RESULTS

The procedure described was used for observations of the first, second, and third Soviet space rockets (launched on 2 January, 12 September, and 4 October 1959, respectively). During the time of the observations, the angle coordinates* of the containers with the scientific apparatus were measured, as were the intensities of the received signal and the character of its variation. Up to now we have determined the parameters of the encounter of the container of the second Soviet space rocket with the moon.

---

*The antenna tracking and the determination of the number of the interference-pattern lobe n, were carried out on the basis of target indications obtained from the coordination-computing center.

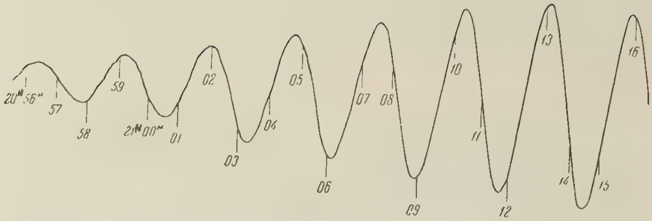


Fig. 5

Fig. 5. Sample of alignment record made on 13 September 1959, using the A-Cygni discrete radio emission source.

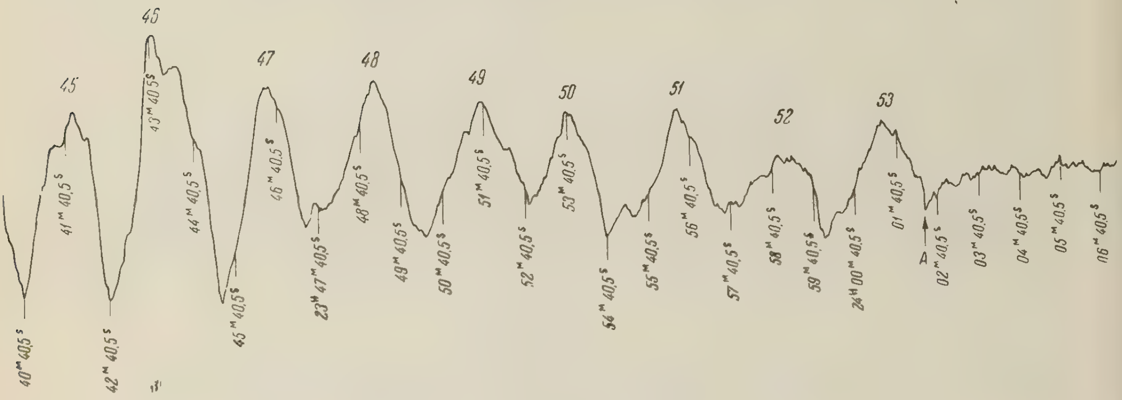


Fig. 6

Fig. 6. Record of signal on the portion where the second Soviet space rocket encountered the moon's surface. The instant of cessation of the signal is denoted by the letter A.

Reliable reception of the signal during the approach to the moon, up to the instant of encounter, made it possible to determine reliably the time of encounter of the rocket with the moon's surface. A copy of the interference pattern of the signal, obtained on a finite portion of the trajectory in one of the receiving channels, is shown in Fig. 6. At the instant when the signal ceased, the sinusoidal character of the record, due to the interference, was replaced by an exponential decay, due to the cessation of the signal and the presence of a time constant. This transition is quite clearly seen (point A on Fig. 6), and determines the time of cessation of the signal as being zero hours 0.2 minutes 22 sec ± 1 sec (14 September 1959).

Taking into account the time of propagation of the signal from the moon to the earth, which amounts to 1.2 sec, this indicates that the encounter between the container and the moon surface occurred at zero hours 0.2 minutes 21 sec ± 1 sec.

The time of cessation of the signal (0 hours 2 minutes 22 sec) corresponds to a lobe number n = 53.43. For this value of n, the curve A(z, n) as calculated by formula (10) crosses the visible disk of the moon.

On recalculating to selenographic coordinates and taking into account the measurement error (which is ± 1' when several readings are averaged), this line is transformed into the region shown shaded in Fig. 7.

Taking into account the data obtained by the automatized measuring set of apparatus for radio control of the orbit [7], it can be assumed that the region where the container with the scientific apparatus landed on the moon is the one shown cross hatched in Fig. 7. The selenographic coordinates of the center of this region are as follows: latitude +30°, longitude -3° (Archimedes crater).

The same figure shows the results of measurements obtained by optical and radio astronomic methods by other authors [8-11].

The intensity of the received signal was determined by comparison with the well known radiation from the discrete source A-Cygni. Simultaneous reception in two polarizations has made it possible to obtain data on the character of variation of the orientation of the container with the scientific apparatus. Fig. 8 shows samples of the record of the signal obtained by the automatic interplanetary station of the third Soviet space rocket. Several characteristic periods of variation of the intensity of the intensity of the signal were observed. Detailed results of these measurements are cited in [12].

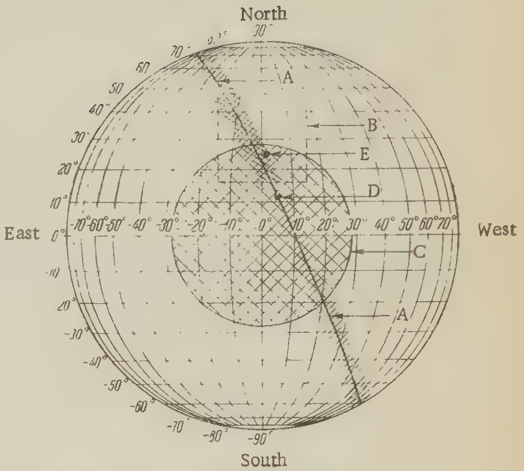


Fig. 7. Landing of container on the moon:

A—data of interferometer of the Physics Institute of the Academy of Sciences USSR; B—data of automatized set [7]; C—data of reference [8]; D—data of [9, 10]; E—data of B. Balazs, J. Balazs and M. Lovas, according to [11]. Single shading — error of radio interferometer; close double shading — proposed region of landing on the moon.

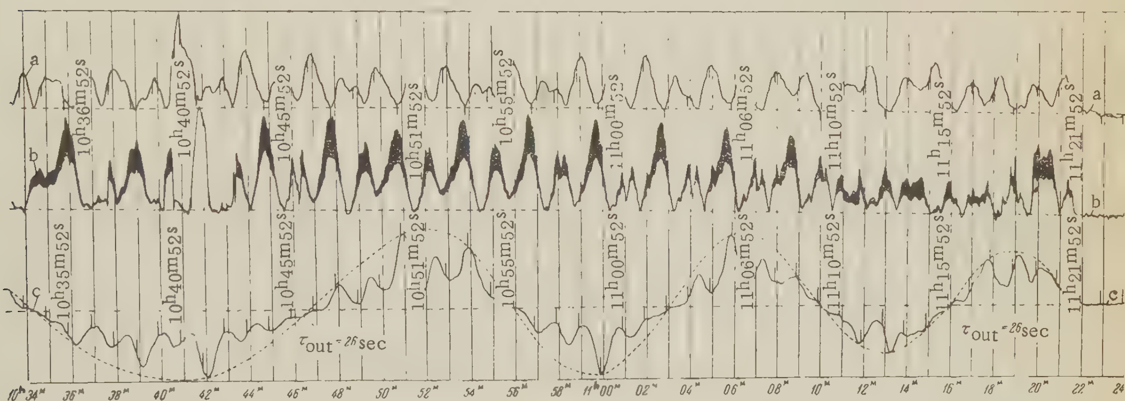


Fig. 8. Samples of recordings of signals made from the third Soviet space rocket on 17 Oct. 195

a—recording of signal obtained from vertical dipole; b — recording of signal obtained from horizontal dipole; c — interference record from horizontal dipoles. Time constant of output unit shif the interference pattern by approximately 15 seconds. If this influence is taken into account the maxima and minima of the signals and records b coincide with those of records c. The maxima of records b and c correspond to the minima of the record a, and vice versa, this indicating rotation of the automatic interplanetary station.

REFERENCES

1. V.V. Vitkevich, Astron. Zhur. (Astronomic Journal), 1952, 29, 4, 450.
2. V.V. Vitkevich, DAN SSSR, 1952, 86, 1, 39.
3. F.V. Bunkin, N.V. Karlov, ZhTF (Journal of Technical Physics) 1955, 25, 4, 793.
4. A.D. Kuz'min, Radiotekhnika, 1957, 12, 1, 12.
5. M. Ryle, Proc. Roy. Soc. A, 1952, 211, 1106, 351.
6. G. Whitfield, Monthly Notices Roy. Astron. Soc., 1957, 117, 6, 680.
7. TASS Report, Newspaper "Pravda" 20 September 1959.
8. J.G. Davies, A.G.B. Lovell, Nature, 1959, 184, 4685, 501.
9. P. Moore, Nature, 1959, 184, 4685, 502.
10. H.P. Wilkins, Nature, 1959, 184, 4685, 502.
11. G. Fielder, Nature, 1960, 185, 4705, 11.
12. V.V. Vitkevich, A.D. Kuz'min, R.L. Sorochenko, V.A. Udal'tsev, Collection of Articles (Artificial Earth Satellites), Academy of Sciences Press, 1961, No. 7.

P.N. Lebedev Physics Institute
Academy of Sciences USSR

Received by editor 4 October 1960

CHARACTER OF FLUCTUATIONS OF A UHF RADIO SIGNAL PROPAGATING OVER AN INHOMOGENEOUS SURFACE

N.D. Filipp

Approximate expressions are obtained for the mean level of the radio signal and its relative fluctuations at the output of a square-law detector for propagation in an inhomogeneous medium over an irregular interface. The values of the normalized functions of the spatial and temporal correlations of the fluctuations of the wave field in the presence of an irregular interface are compared with the values of these functions in free space.

The deductions obtained are compared with the experimental data.

1. ESTIMATE OF UHF FLUCTUATIONS IN THE CASE OF MULTIPATH PROPAGATION

References [1, 2, 3] report on investigations of the fluctuation characteristics of electromagnetic waves with account of the influence of the interface. The interface was assumed in these investigations to be a perfectly conducting plane. When the problem is so formulated, it is impossible to explain the sharp differences between the mean levels of the electric field at points located in horizontal planes transverse to the path. To explain these differences, account must be taken of the irregularities of the interface.

Experiments which we have carried out to investigate the fluctuation characteristics of waves in the 10-cm band on a short path (37 km) in the direct visibility zone, using reception at separated points, disclosed that the distribution of the field transverse to the path is quite complicated (even within the limits of several wavelengths).

We consider below, in the geometrical-optics approximation [4], the effect of a rough interface on the mean level of the received signal and on its absolute and relative fluctuations.

If the transmitter is located at point A and the receiver at point P (Fig. 1), then the field $\mathscr{E}$ (P, t) at the point of reception is determined by the field of the direct ray $\mathscr{E}_1$ (P, t) and by the field of the reflected ray $\mathscr{E}_2$ (P, t)

where
$$\mathscr{E}(P,t) = \mathscr{E}_1(P,t) + \mathscr{E}_2(P,t),$$
$$\mathscr{E}_1(P,t) = [E_{01}(P,t) + \Delta E_1(P,t)] \exp\{i[\varphi_{01}(P,t) + \Delta\varphi_1(P,t) - \omega t]\};$$
$$\mathscr{E}_2(P,t) = R(C)[E_{02}(P,t) + \Delta E_2(P,t)] \exp\{i[\varphi_{02}(P,t) + \Delta\varphi_2(P,t) - \omega t]\}. \tag{1}$$

To simplify the notation we shall henceforth leave out the arguments P, and t. The quantities E_{01}, E_{02}, φ_{01}, and φ_{02} are the mean amplitudes and phases of the fields of the direct and reflected rays; ΔE_1, ΔE_2, $\Delta\varphi_1$ and $\Delta\varphi_2$ are the fluctuating components of the corresponding quantities, and φ_{02} contains also the phase loss angle due to reflection; R(C) is the absolute value of the reflection coefficient.

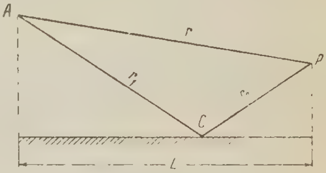


Fig. 1. Schematic representation of double-beam model of propagation.

The absolute value of the reflection coefficient depends not only on the electric properties of the earth, but also on the geometry of the reflecting surface. Thus, in the case of convex inhomogeneities, rays which were nearly parallel prior to reflection, become divergent after reflection and consequently the amplitude of the reflected wave decreases with distance. The degree of divergence depends on the curvature of the surface at the point of reflection. In the case of concave inhomogeneities, the rays can become concentrated at certain points in space, and consequently, an increase at the point of reception is possible compared with the amplitude of the point C.

If we confine ourselves to convex inhomogeneities then, as follows from [5], the absolute value of the reflection coefficient can be represented in the form

$$R(C) = R_0 \sqrt{\frac{D(0)}{D(r_2)}}, \tag{2}$$

where R_0 is the coefficient of reflection from a perfectly conducting plane;

$$D(r_2) = \cos\theta + 2r_2\left[\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right)\cos^2\theta + \frac{\sin\theta}{\rho_0}\right] + \frac{2r_2^2}{\rho_1\rho_2}\cos\theta;$$

$D(r_2)dS$ is the area of the normal cross section of the beam incident on the surface element dS, containing the point of reflection; θ is the angle of incidence; ρ_1 and ρ_2 are the principal radii of curvature of the normal cross section of the surface; $K = 1/\rho_1\rho_2$ is the Gaussian curvature of the surface; ρ_0 is the radius of the normal intersection between the surface and the plane of incidence of the beam. The quantity $D(r_2)/D(0)$ obviously yields the broadening of the beam, i.e., the ratio of its cross section at a distance r_2 from the surface (reckoned along the ray) to the cross section at the surface itself.

At small values of geometric path difference, we have $E_{01} = E_{02}$ [10]. Let us consider the case of small fluctuations, $\Delta E_i/E_{01} \ll 1$ and $\Delta\varphi_i \ll 1$, which is valid in the direct visibility zone. Assuming the earth to be perfectly conducting and leaving out the time factor in the expressions (1), we obtain for the mean value of the signal at the output of a square-law detector

$$\overline{U(P,t)} = \alpha\,[\overline{(E_{01} + \Delta E_1)^2} + R^2(C)\,\overline{(E_{01} + \Delta E_2)^2} + \\ + 2R(C)\,\overline{(E_{01} + \Delta E_1)(E_{01} + \Delta E_2)\cos\Delta\varphi}], \tag{3}$$

where $\Delta\varphi = \Delta\varphi_0 + \Phi = \varphi_{02} - \varphi_{01} + \Delta\varphi_2 - \Delta\varphi_1$; α is a proportionality coefficient, which can be readily determined during the process of calibration. The superior bar denotes statistical averaging.

It is assumed that the fluctuations of the amplitude and of phase are independent, as takes place, for example, when the scattering is by turbulent inhomogeneities and when the wave parameter D is sufficiently large [6]. Making use of the fact that the phase of a wave scattered by random inhomogeneities of the refractive index has a normal distribution [7], we find that the phase difference $\Phi = \Delta\varphi_2 - \Delta\varphi_1$ also has a normal distribution [8], and then

$$\overline{\cos(\Delta\varphi_0 + \Phi)} = \operatorname{Re}\left\{\frac{1}{\sqrt{2\pi}\,\sigma_\varphi}\int_{-\infty}^{\infty} e^{i(\Delta\varphi_0+\Phi)}e^{-\frac{\Phi-\Phi_0}{2\sigma_\varphi^2}}\,d\Phi\right\} = \cos(\Delta\varphi_0 + \Phi_0)\,e^{-\frac{\sigma_\varphi^2}{2}},$$

where

$$\Phi_0 = \bar{\Phi} = \overline{\Delta\varphi_2} - \overline{\Delta\varphi_1};\ \ \sigma_\varphi^2 = \sigma_1^2 - 2f_\varphi + \sigma_2^2;\ \ \sigma_1^2 = \overline{(\Delta\varphi_2 - \overline{\Delta\varphi_1})^2};$$

$$\sigma_2^2 = \overline{(\Delta\varphi_2 - \overline{\Delta\varphi_2})^2};\ \ f_\varphi = \overline{(\Delta\varphi_1 - \overline{\Delta\varphi_1})(\Delta\varphi_2 - \overline{\Delta\varphi_2})}.$$

If the irregular components of the fields are such that $\overline{\Delta E_i} = \overline{\Delta\varphi_i} = 0$ (i.e., ΔE_i and $\Delta\varphi_i$ have a symmetrical distribution [9]), then (3) can be reduced to the form

$$\begin{aligned}\overline{U(P, t)} = \alpha\{&\overline{\Delta E_1^2} + R^2(C)\,\overline{\Delta E_2^2} + 2R(C)\,\overline{\Delta E_1 \Delta E_2}\cos\Delta\varphi_0 e^{-\frac{\sigma_\varphi^2}{2}} + \\ &+ E_{01}^2[1 + R^2(C)] + 2R(C)\,E_{01}^2\cos\Delta\varphi_0 e^{-\frac{\sigma_\varphi^2}{2}}\}.\end{aligned} \tag{4}$$

At small fluctuations of the wave field (which is valid in the direct-visibility zone when the paths are not very long), when $e^{-\frac{\sigma_\varphi^2}{2}} \simeq 1 - \frac{\sigma_\varphi^2}{2}$ and $\overline{E} \simeq E_{01}\sqrt{1 + R^2(C) + 2R(C)\cos\Delta\varphi_0}$, we obtain for the mean value of the fluctuations of the signal power

$$\begin{aligned}\overline{V(P, t)} &= \overline{U(P, t)} - \alpha\{E_{01}[1 + R^2(C) + 2R(C)\cos\Delta\varphi_0]\} = \\ &= \alpha\{\overline{\Delta E_1^2}(1 + R^2(C) + 2R(C)\,\varrho_{\Delta E}\cos\Delta\varphi_0) - 2R(C)\,\overline{\Delta\varphi_1^2}(1 - \varrho_{\Delta\varphi})\times \\ &\times(E_{01}^2 + \overline{\Delta E_1 \Delta E_2})\cos\Delta\varphi_0\},\end{aligned} \tag{5}$$

where $\varrho_{\Delta E} = \overline{\Delta E_1 \Delta E_2}/\overline{\Delta E_1^2}$ and $\varrho_{\Delta\varphi} = \overline{\Delta\varphi_1 \Delta\varphi_2}/\overline{\Delta\varphi_1^2}$ are the coefficients of spatial correlations of the fluctuations of the amplitudes and phases of the direct and reflected waves in the aperture of the receiving antenna. It is proposed in [3] that $\overline{\Delta E_1^2} \simeq \overline{\Delta E_2^2} \simeq \sigma_E^2$ and $\overline{\Delta\varphi_1^2} = \overline{\Delta\varphi_2^2} = \sigma_1^2$. From (4) and (5) it is seen that both the mean level of the signal at the output of the system, and its fluctuations depend on the difference in the path of the direct and reflected rays (which causes their distribution in space to have a lobe-like character), and also on the absolute value of the coefficient of reflection.

The fluctuations of the signal at the output of the system are determined not only by the fluctuations of the amplitudes and phases of the field, but also by the degree of correlation of the fluctuations of the amplitudes ΔE_i and of the phases $\Delta\varphi_i$ ($i = 1, 2$). Expression (5) in the vicinity of the interference minima and maxima of the mean level of the field are reduced to the form

$$\overline{V(P, t)} = \alpha\{\sigma_E^2[1 + R^2(C) \mp 2R(C)\,\varrho_{\Delta E}] \pm 2R(C)\,E_{01}^2\sigma_1^2(1 - \varrho_{\Delta\varphi})\}. \tag{6}$$

If the fluctuations of the wave field of the direct and reflected rays are fully correlated, i.e., $\varrho_{\Delta E} = \varrho_{\Delta\varphi} = 1$, then

$$\begin{aligned}&\overline{V(P, t)} = \alpha\{\sigma_E^2[1 + R(C)]^2\} && \text{in the vicinity of the maxima, and} \\ &\overline{V(P, t)} = \alpha\{\sigma_E^2[1 - R(C)]^2\} && \text{in the vicinity of the minima.}\end{aligned} \tag{7}$$

On the other hand, if the fluctuations of the directed and reflected rays do not correlate, we have

$$\overline{V(P,t)} = \alpha\{\sigma_E^2[1+R^2(C)] - 2R(C)E_{01}^2\sigma_1^2\} \qquad \text{in the vicinity of the maxima, and} \tag{8}$$

$$\overline{V(P,t)} = \alpha\{\sigma_E^2[1+R^2(C)] + 2R(C)E_{01}^2\sigma_1^2\} \qquad \text{in the vicinity of the minima.} \tag{9}$$

For the relative fluctuations of the signal level we obtain

$$\frac{\overline{V(P,t)}}{\overline{U(P,t)}} \simeq$$

$$\simeq \frac{\sigma_E^2\left[1+R^2(C)-2R(C)\rho_{\Delta E}\cos\left(2\pi\frac{\Delta r}{\lambda}\right)\right]+2R\sigma_1^2(1-\rho_{\bar{\Delta}\varphi})(E_{01}^2\overline{+\Delta E_1\Delta E_2})\cos\left(2\pi\frac{\Delta r}{\lambda}\right)}{E_{01}^2\left[1+R^2(C)-2R(C)\cos\left(2\pi\frac{\Delta r}{\lambda}\right)\right]},$$

which displays the lobe-like character of the relative fluctuations in the presence of a reflected component.

2. COEFFICIENT OF SPATIAL CORRELATION

Let us examine in the approximation of geometrical optics [4] a two-beam propagation model, i.e., let us assume that two beams arrive at the receiving antenna, one direct and one reflected. If receiving antennas are located at the points P_1 and P_2, then the resultants of the direct and reflected fields in the apertures of the antennas have the following form

$$\mathscr{E}(P_1) = \mathscr{E}_1(P_1) + \mathscr{E}_1'(P_1),$$
$$\mathscr{E}(P_2) = \mathscr{E}_2(P_2) + \mathscr{E}_2'(P_2),$$

where

$$\begin{aligned}
\mathscr{E}_1(P_1) &= [E_{01}(P_1,t) + \Delta E_1(P_1,t)]\exp\{i[\varphi_{01}(P_1,t) + \Delta\varphi_1(P_1,t) - \omega t]\},\\
\mathscr{E}_1'(P_1) &= R_1(C_1)[E_{01}'(P_1,t) + \Delta E_1'(P_1,t)]\exp\{i[\varphi_{01}'(P_1,t) +\\
&\quad + \Delta\varphi_1'(P_1,t) - \omega t]\},\\
\mathscr{E}_2(P_2) &= [E_{02}(P_2,t) + \Delta E_2(P_2,t)]\exp\{i[\varphi_{02}(P_2,t) + \Delta\varphi_2(P_2,t) - \omega t]\},\\
\mathscr{E}_2'(P_2) &= R_2(C_2)[E_{02}'(P_2,t) + \Delta E_2'(P_2,t)]\exp\{i[\varphi_{02}'(P_2,t) +\\
&\quad + \Delta\varphi_2'(P_2,t) - \omega t]\}.
\end{aligned} \tag{10}$$

The coefficient of spatial correlation of the fluctuations of the envelopes of the signal amplitudes at the output of square-law detectors

$$\rho(P_1,P_2) = \frac{\overline{U(P_1)U(P_2)} - \overline{U(P_1)}\,\overline{U(P_2)}}{\{[\overline{U^2(P_1)} - (\overline{U(P_1)})^2][\overline{U^2(P_2)} - (\overline{U(P_2)})^2]\}^{1/2}},$$

where $U(P_1) = \alpha\mathscr{E}(P_1)\ \mathscr{E}^*(P_1)$ and $U(P_2) = \alpha\mathscr{E}(P_2)\ \mathscr{E}^*(P_2)$ have in the general case, as follows from (10), a rather complicated form, but to describe their spatial distribution connected with the lobe-like character of the interference field we can use the approximate expressions obtained in some particular cases.

If both receiving antennas are located in the vicinities of the interference maxima, or minima of the field (with the exception of the minima for which $R_1 \simeq R_2 \simeq 1$), then under the following conditions:

a) $\Delta E_i/E_{01} \ll 1$, $\Delta\varphi_i \ll 1$;
b) ΔE_i and $\Delta\varphi_i$ have a normal distribution;
c) $D = 4L/Ka^2 \gg 1$;
d) the direct and reflected waves are weakly correlated;
e) the character of the fluctuations (depth and frequency) of the reflected waves is the

same as in the free space, the coefficient of spatial correlation of the signal fluctuations assumes the following approximate form

$$\rho(P_1, P_2) = \frac{\overline{\Delta E_1 \Delta E_2}(1 \pm R_1 \pm R_2 + R_1R_2) + \overline{\Delta E_1' \Delta E_2'}(R_1^2R_2^2 \pm R_1R_2^2 \pm R_1^2R_2 + R_1R_2)}{[\overline{\Delta E_1^2}(1 \pm R_1)^2(1 + R_1^2)\overline{\Delta E_2^2}(1 \pm R_2)^2(1 + R_2^2)]^{1/2}}. \tag{11}$$

On the other hand, if one of the antennas is located in the vicinity of the maximum of the interference field (assume that this is the first antenna), and the other in the vicinity of the minimum field, we obtain

$$\rho(P_1, P_2) = \frac{\overline{\Delta E_1 \Delta E_2}(1 + R_1 - R_2 - R_1R_2) + \overline{\Delta E_1' \Delta E_2'}(R_1^2R_2^2 + R_1R_2^2 - R_1^2R_2 - R_1R_2)}{[\overline{\Delta E_1^2}(1 + R_1)^2(1 + R_1^2)\overline{\Delta E_2^2}(1 - R_2)^2(1 + R_2^2)]^{1/2}}. \tag{12}$$

In this case, the coefficient of spatial correlation, depending on the values of $\overline{\Delta E'_1 \Delta E'_2}$, R_1 and R_2 can be considerably less than in case (11).

It follows from (11) and (12) that the spatial correlation of the fluctuations of the amplitudes of the received signals has a lobe-like character in the presence of an interface.

In particular, if the correlation of the fluctuations of the reflected waves is the same as of the direct waves, i.e., $\overline{\Delta E_1 \Delta E_2} = \overline{\Delta E'_1 \Delta E'_2}$ and $R_1 \simeq R_2 = R$, then expressions (11) and (12) assume the following respective forms

$$\varrho(P_1, P_2) = \frac{\overline{\Delta E_1 \Delta E_2}}{\sqrt{\overline{\Delta E_1^2}\,\overline{\Delta E_2^2}}} = \rho_f(P_1, P_2), \tag{13}$$

$$\varrho(P_1, P_2) = \frac{\overline{\Delta E_1 \Delta E_2}(1 - R^2)}{\sqrt{\overline{\Delta E_1^2}\,\overline{\Delta E_2^2}(1 + R^2)}} = \varrho_f(P_1, P_2)\frac{1 - R^2}{1 + R^2} < \varrho_f(P_1, P_2), \tag{14}$$

where $\varrho_f(P_1, P_2)$ is the coefficient of correlation of the fluctuations of the field amplitudes at the points P_1 and P_2 in free space.

The decorrelation of the received signals can occur also as a result of a change in the character and the degree of correlation of the waves reflected from the underlying surface. Thus, in points of space where the reflected waves are weakly correlated, (11) assumes the form

$$\rho(P_1, P_2) = \rho_f(P_1, P_2)\frac{1 + R_1 + R_2 + R_1R_2}{[(1 \pm R_1)^2(1 + R_1^2)(1 \pm R_2)^2(1 + R_2^2)]^{1/2}}. \tag{15}$$

Unlike the case (13), the correlation coefficient under such conditions can prove to be also much less than expected in free space. In particular, when $R_1 = R_2 = R$

$$\rho(P_1, P_2) = \frac{\rho_f(P_1, P_2)}{1 + R^2} < \rho_f(P_1, P_2). \tag{16}$$

When the antennas are located in the vicinities of the field minima, wherein $R_1 \simeq R_2 \simeq 1$, we obtain

$$\rho(P_1,P_2) = \frac{(\overline{\Delta E_1^2 \Delta E_2^2} - \overline{\Delta E_1^2}\,\overline{\Delta E_2^2}) + (\overline{\Delta E_1'^2 \Delta E_2'^2} - \overline{\Delta E_1'^2}\,\overline{\Delta E_2'^2})}{\{[\overline{\Delta E_1^4} - (\overline{\Delta E_1^2})^2 + \overline{\Delta E_1'^4} - (\overline{\Delta E_1'^2})^2][\overline{\Delta E_2^4} - (\overline{\Delta E_2^2})^2 + \overline{\Delta E_2'^4} - (\overline{\Delta E_2'^2})^2]\}^{1/2}}. \tag{17}$$

The behavior of the correlation coefficient of the signal fluctuations at the output of the system differs appreciably in this case, above all because it is expressed in terms of correlation functions of the squares of the fluctuating amplitudes of the field. Thus, it follows from (11)—(17) that in the case of double-beam propagation of radio waves, the space correlation of the fluctuations of the received signals has a lobe-like character and can be under various conditions quite less than its value in free space under the same atmospheric conditions. In the case of multiple-beam propagation and weak interrelation between the fluctuations of the reflected waves, this difference can be even greater. This is physically clear, for the greater the number of statistically independent waves arriving at the points P_1 and P_2, the weaker the statistical connection between the resultant fields at these points.

3. TEMPORAL AUTOCORRELATION COEFFICIENT

If a receiving antenna is located at the point P and if we denote by U(P, t) and U(P, t + τ) the signals at the output of a square-law detector at the instants of time t and t + τ, where τ is the time shift, then we can neglect, in the vicinity of the interference maxima and minima (with the exception of the minimum points where R $\simeq$ 1), as in preceding particular cases (11)—(16), the terms containing moments of fourth order. In this particular case, recognizing that under the assumptions made we have $\overline{\Delta E_1 \Delta E_{1\tau}} \simeq \overline{\Delta E_2 \Delta E_{2\tau}}$ where ΔE_1, $\Delta E_{1\tau}$, ΔE_2, $\Delta E_{2\tau}$ — the amplitudes of the fluctuating components of the field of the direct and reflected waves at the instants t and t + τ, we obtain

$$\rho(P, \tau) = \frac{\overline{U(P, t)\, U(P, t+\tau)} - \overline{U(P, t)}\; \overline{U(P, t+\tau)}}{\{[\overline{U^2(P, t)} - (\overline{U(P, t)})^2]\, [\overline{U^2(P, t+\tau)} - (\overline{U(P, t+\tau)})^2]\}^{1/2}} =$$

$$= \frac{\overline{\Delta E_1 \Delta E_{1\tau}}\, [(1 \pm R)^2 + (R^2 \pm R)^2]}{\overline{\Delta E_1^2}\, [(1 \pm R)^2 + (R^2 \pm R)^2]} = \frac{\overline{\Delta E_1 \Delta E_{1\tau}}}{\overline{\Delta E_1^2}} = \rho_{\mathbf{f}}(\tau). \tag{18}$$

We denote here by ϱ (P, τ) the coefficient of temporal autocorrelation of the fluctuations of the envelope of the signal amplitude at the output of the quadratic system at the point P, in the presence of an interface, and by $\varrho_f(\tau)$ the coefficient of temporal autocorrelation in free space. Thus, in this case the spatial distribution of the temporal correlation of the signal-amplitude fluctuations at the output of the receiving signal does not have a lobe-like character and is equal to the correlation of the fluctuations of the field amplitude in free space.

Naturally, if we take into account the dynamic nature of the underlying interface, the temporal correlation of the fluctuations of the received signal may prove to be different than in free space.

At the interference minima, when R $\simeq$ 1, we can neglect the terms containing moments of second order; we then obtain for the coefficient of temporal correlation of the fluctuations of the signal amplitude at the output of the receiving system

$$\rho(P, \tau) = \frac{\overline{\Delta E_1^2 \Delta E_{1\tau}^2} - \overline{\Delta E_1^2}\; \overline{\Delta E_{1\tau}^2}}{\overline{\Delta E_1^4} - \left(\overline{\Delta E_1^2}\right)^2} = \rho'(\tau) = \rho^2_{\mathbf{f}}(\tau), \tag{19}$$

where $\varrho'(\tau)$ denotes the coefficient of temporal autocorrelation of the fluctuations of the square of the field amplitudes at the input of the system, which in our case, where we assume normal distribution of the scattered field components, is equal to the square of the coefficient of temporal correlation of the fluctuations of the field amplitudes at the input of the system.

4. DISCUSSION OF EXPERIMENTAL RESULTS

It is clear from the foregoing that fluctuations at the output of a receiving signal, in the presence of an inhomogeneous interface, are different at different points of space and deviate from the values expected in free space. However, the degree wherein they differ from the values in free space depends, on the one hand, on the magnitude and character of the reflecte waves, as determined by the reflection coefficient R, and on the other hand it depends on the state of the troposphere. The value of R will be determined essentially by the character of the path, by the form of inhomogeneities of the interface, by its electrical properties, and b the vegetation cover.

We give below certain results of experimental researches, carried out in 1959-1960 on an open dry path 37 km long (Fig. 2). The character of the path gave grounds for assuming that the coefficient of reflection is insignificant, with the exception of the locality near the receiving point. The essentially reflecting zone, located near the receiving point, was a faintly uneven surface. The smallest clearance, allowing for normal reflection, amounted t

15-20 meters, while the maximum deviation of the profile from a sphere was about 50 meters.

On one end of the path was placed a pulsed transmitter with pulse power of 80 kw, operating frequency $f = 3,000$ Mc, a pulse repetition rate of 400/sec, and a pulse duration of 1 microsecond.

The radiating system was a parabolic antenna 1.8 meters in diameter, radiating a horizontally polarized wave.

The receiving apparatus, located on the other end of the path, had two identical channels. Each channel was a superheterodyne receiver with apparatus which permitted automatic synchronous registration of the levels of the signals received by two spatially-separated antennas with variable base. The receiving antennas used were symmetrical dipoles 4 meters above the earth.

The statistical characteristics of the received signals were estimated (see sections 1-3) at the outputs of the square-law detectors, since the power amplitude characteristics of the receivers were linear.

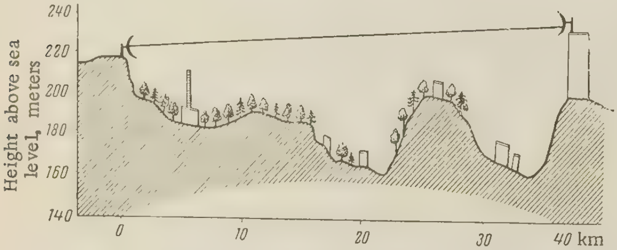


Fig. 2. Profile of path.

We investigated the transverse correlation of the field amplitude fluctuations in a horizontal plane. During the time of the experiment, the stability of the receiving and transmitting systems were continuously monitored, for which purpose part of the radiated power was diverted to the power meter of the transmitter, and to a control standard signal generator. A block diagram of the receiving apparatus is shown in Fig. 3.

The measurements were carried out three times a day with the sessions lasting from 30 to 80 minutes. The spacing between antennas was varied between one and 40 wavelengths. Continuous recording of the envelope of the received-signal amplitudes was made on motion picture film for 5 - 7 minutes. The rate of travel of the film depended on the character of the fluctuations and ranged from 1 to 10 mm/sec. The time interval for processing was 3 - 5 minutes which, on the one hand, excluded the influence of the non-stationary nature of the investigated processes, and on the other hand permitted a sufficiently accurate determination of such statistical characteristics of the field variation as, for example, the mean value, the quadratic deviation, the correlation coefficients, etc. [11].

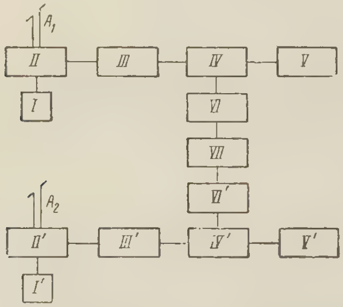


Fig. 3. Block diagram of receiving system. I, I'—heterodyne; II, II'—mixers with preamplification of intermediate frequency; III, III'—IF amplifier; IV, IV'—video amplifier; V, V'—cathode ray oscillographs for visual observation; VI, VI'—peak detectors with DC amplifiers; VII—loop oscillograph; A_1A_2—antennas.

The experimental data were used to calculate the following: a) mean value of the signal at the output of the system; b) mean-square deviation of the signal; c) normalized temporal autocorrelation functions; d) normalized spatial correlation functions.

To determine the transverse correlation fluctuations of the field amplitude we used the method of successive recordings of the signal at various points of space using a single stationary antenna, a method which is fully applicable if the statistical state of the troposphere is stationary.

Since the experiment extended over many months (June 1959 - June 1960), we were able to investigate the fluctuating characteristics of the wave field at different values of the reflected field component. Depending on the electric properties, the character of the unevenness, and the vegetation on the reflecting surface, the effective reflection coefficient had a range $0.9 \lesssim R \lesssim 0.1$.

The results of the investigation of the fluctuation characteristics of the received radio signal confirm the conclusions that when the mean field level has a lobe-like structure the relative fluctuations of the field amplitude and their spatial correlation also have a lobe-like character. No dependence of the distribution of the temporal correlation of the field-amplitude fluctuations on the interference structure of the mean field level was observed.

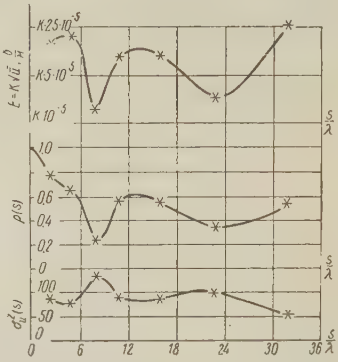


Fig. 4. Graphic representation of mean field level E, coefficient of space-transverse correlation $\rho(s)$, and the dispersion $\sigma_u^2(s)$ of the received signal; s — base between the receiving antenna; *— experimental values

Fig. 4 shows values of the average level of the field E, of the coefficient of transverse correlation ϱ (s) and the dispersion $\sigma_u^2(s)$ of the received radio signal in one of the measurement sessions. We note that the dispersion $\sigma_u^2(s)$ (Fig. 4) is obtained with the mean value of the signal at the output of the receiver reduced to one and the same level, this being accomplished by setting the gain of the receiver, so that essentially $\sigma_u^2(s)$ is a quantity proportional to the relative magnitude of the fluctuations. The course of the spatial correlation of the amplitude fluctuations of the received signal, shown in Fig. 4 can be readily explained on the basis of (11) and (12), namely that when the moving antenna is located near the interference maximum of the field and the stationary antenna is in the vicinity of the minimum, the factor $(R_1^2R_2^2 + R_1R_2^2 - R_1^2R_2 - R_1R_2)$ of the correlation function of the amplitudes of the reflected waves $\overline{\Delta E'_1 \Delta E'_2}$ becomes negative, whereas when the moving antenna is in the vicinity of the maxima the factor $(R_1^2R_2^2 + R_1R_2^2 + R_1^2R_2 + R_1R_2)$ of the function $\overline{\Delta E'_1 \Delta E'_2}$ is a positive quantity, so that stronger decorrelation of the received signals takes place for the interference minima of the mean level of the field.

Since the correlation function $\Delta E_1 \Delta E'_2$ decreases with increasing separation base, it follows from (12) that the difference between the neighboring maximal values of the coefficient of transverse correlation $\rho_{max} - \rho_{min}$ decreases with increasing base. This effect is clearly pronounced in all experiments where the mean level of the field has a lobe-like distribution (Figs. 4 and 6). Fig. 5 shows the experimental data for the coefficient of temporal autocorrelation, obtained for different points in space at which the mean field level, the relative fluctuations, and their dispersion greatly differed (the data cited pertain to the processing of recordings made of the same measurement session as for the spatial correlation function shown in Fig. 6a). It is seen from Fig. 5 that within the framework of the permissible data-reduction errors, these do not differ greatly from each other and are well approximated by the exponential function $e^{-\frac{v_\perp \tau}{a}}$ (solid curve), where $v_\perp$ is the drift-velocity component normal to the vertical plane passing through the receiver and transmitter, τ is the time shift, and a is a parameter characterizing the mean dimension of the inhomogeneities of the medium, responsible for the fluctuations of the wave field.

The degree of deviation of the relative fluctuations and the spatial correlation of the amplitudes of the signals in different reception points depend on the depth of the interference pattern of the mean field level. Fig. 6 shows the behavior of the transverse correlation for different values of the effective reflection coefficient R_e. The reflection coefficient was determined from the vertical distribution of the mean field level [12]:

$$R_{\partial} = \frac{E_{max} - E_{min}}{E_{max} + E_{min}}$$

We note that the depth of the interference pattern in the horizontal plane, transverse to the path, was almost the same as for a vertical plane, and differs only in having a more complicated field distribution as a function of the character of the irregularities of the interface.

In the case of a field with a deep interference pattern (Fig. 6a), the distribution of the transverse correlation of the fluctuations of the amplitude of the envelope is quite complicated, and its character is clearly connected with the lobe-like distribution of the interference field. However, when $R_{eq} \leqslant 0.3$, the lobe character of the distribution of the spatial correlation and of the relative fluctuation of the field amplitude almost disappears.

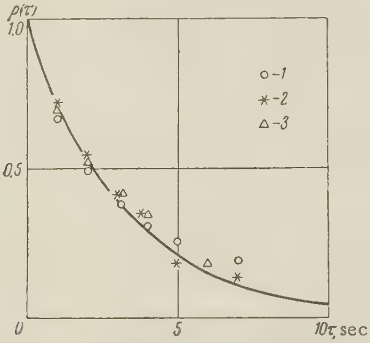


Fig. 5. Graphic representation of temporal coefficient of autocorrelation; 1, 2, 3—experimental values at different points of space.

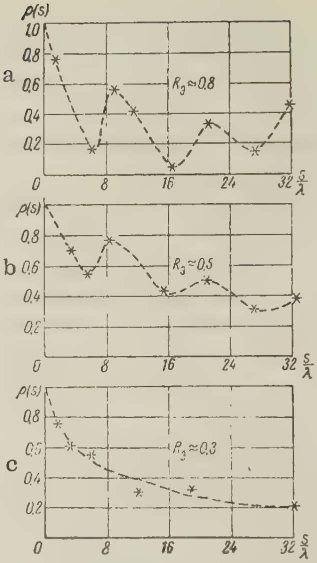


Fig. 6. Graphic representation of the coefficient of transverse correlation for different values of the effective reflection coefficient.

Whereas in the former case the lobe-like structure of the fluctuation characteristics of the field can mask certain effects due to atmospheric inhomogeneities (in particular, correlation distance), in the case of a weak coherent field components ($R_e \leqslant 0.3$) it is possible to make a complete study of the behavior of the fluctuation characteristics of the field as a function of the state of the troposphere.

We note that the presence of stationary interfaces, as follows from the foregoing, cannot give rise to field fluctuations. It leads only to a unique distribution of these characteristics in space.

I take this opportunity to thank A.A. Semenov and G.A. Karpeyev for valuable help rendered in the performance of this work.

REFERENCES

1. A.V. Men', V.I. Gorbach, S.Ya. Braude. News of the Universities, Ministry of Higher Education USSR — Radiophysics, 1959, 2, 3, 388.
2. E.A. Kaner, F.G. Bass, ibid, 2, 4, 553.
3. F.G. Bass, E.A. Kaner, ibid, 2, 4, 565.
4. B.A. Vvedenskiy, A.G. Arenberg. Problems in the propagation of UHF waves, Part 1. Soviet Radio Press, 1948.
5. V.A. Fok, Uspekhi Fiz. Nauk 1948, 36, 3, 308.
6. L.A. Chernov. Propagation of Waves in a Medium with Random Inhomogeneities. Academy of Sciences Press, USSR, 1958.
7. R.B. Muchmore, A.D. Wheelon. Proc. IRE, 1955, 43, 10, 1437.
8. B.V. Gnedenko. Course in Probability Theory, 1950.
9. G.A. Karpeyev, A.A. Semenov, Radiotekhnika i elektronika, 1950, 5, 4, 578.
10. M.P. Kolukhanov. Propagation of Radio waves. Svyaz'izdat, 1951.
11. A.A. Semenov, G.A. Karpeyev, News of the Moscow University, 1958, 4, 71.
12. A.G. Arenberg. Propagation of Decimeter and Centimeter waves. Soviet Radio Press, 1957.

Physics Faculty, M.V. Lomonosov Moscow State
University, Chair of Radio Wave Propagation

Received by editor 30 September 1960.

CERTAIN OPTIMAL DEVICES FOR DETECTION OF A PULSE SIGNAL OF RANDOM DURATION IN THE PRESENCE OF NOISE

N.K. Kul'man, R.L. Stratonovich

We describe a theoretical investigation of the operation of optimal nonlinear and linear filters with relays used as the decision element for the separation of an asymmetrical pulse signal of random duration from white noise. Theoretical estimates are obtained for the average number of false pulses per unit time and for the probability of failure of detecting pulses, depending on their duration, assuming the noise to be small. We plot the characteristics of detection of a single pulse of random duration for several particular cases. We compare the operation of a nonlinear and linear filter.

INTRODUCTION

The receiver input is usually a signal r(t) which in the simplest case is the sum of a useful signal and of noise: $r(t) = s(t) + n(t)$. The task of the optimal receiver is to produce at its output a signal $\widetilde{s}(t)$ that differs least from the useful signal s(t). A theoretical analysis of the problem of optimum filtration is found in many papers [1-5]. The Kolmogorov-Wiener theory [1, 2] gives the optimal filter in the class of linear transformations, which coincides with the optimal filter itself (on the basis of the mean-square criterion) in the case when the signal and noise have Gaussian distributions. Many of the encountered signals, however, are essentially non-Gaussian, such as pulsed signals of various kinds. In the case of a non-Gaussian signal or noise, the nonlinear transformation can be expected to be certainly better, and the optimal system for this case is a nonlinear filter. To calculate the nonlinear optimal filtering systems one can use successfully the Markov theory of conditional processes [3, 4, 9]. Optimum nonlinear systems are simplest to calculate when the signal or its parameters can assume only a finite number n of possible values. The equation for the a posteriori probabilities of these values, derived in [4], can be directly represented by a model. In the present article we consider the case $n = 2$. It is assumed that the useful signal is a Markov process i.e., the time that the signal remains in each of the two possible states has an exponential a priori distribution law. The received signal is the sum of the described pulse signal and white noise.

If we consider the stationary problem, then the optimum filtering system can also be calculated from the Kolomogorov-Wiener theory. The latter optimal filtering system will certainly be poorer than the nonlinear system designed on the basis of the theory of conditional Markov process, inasmuch as in the Kolmogorov-Wiener theory the optimal system is sought only in the class of linear systems, whereas the optimal system itself is nonlinear.* It is interesting, however, to compare these systems directly and to ascertain the extent by which they differ in operation, and under what conditions this difference is greater or smaller. A theoretical comparison entails great computational difficulties, so that we shall confine ourselves to the case of an asymmetrical signal and also to the case of white noise.

---

*By way of an optimality criterion we can take any of the following: minimum mean-squared error, minimum modulus of error, or minimum summary mean error time. For a signal in two positions, these quantities are proportional to each other.

It was found possible to obtain for these cases approximate estimates of certain quantities that characterize the efficiencies of the nonlinear and linear filters, to compare them with each other and with the experimental results described in [6]. We choose for such quantities the average number of false pulses per unit time and the average relative number of undetected pulses, as functions of their duration. We also considered the case of filtering with finite observation interval and obtained the characteristics of optimal receivers in analytic form.

1. EQUATION OF OPTIMAL NONLINEAR FILTER AND OF LINEAR FILTER CALCULATED BY THE KOLMOGOROV-WIENER THEORY

Let us consider the problem of filtering a generalized telegraph signal, comprising a sequence of rectangular pulses which assume values +a and -a, with a specified average of α transitions per second from the state +a to the state -a, and a specified β transitions per second from the state -a to the state +a. The noise is assumed to be white with spectral density N($\langle nn_\tau \rangle = N\delta(\tau)$). The a priori probabilities w^+ and w^- of being in the states +a and -a satisfy accordingly the following system of equations

$$\dot{w}^+ = -\alpha w^+ + \beta w^-,$$
$$\dot{w}^- = \alpha w^+ - \beta w^-.$$

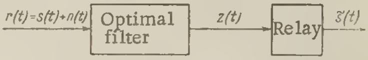


Fig. 1. Diagram of optimal receiver.

The useful signal is thus a Markov process. Because of this circumstance we can find an optimal transformation in an arbitrary class, without confining ourselves to a prescribed class of transformations. Applying Eq. (88) of [4] to the particular case described here, we obtain the following equation for optimum filtering, written in terms of the dimensionless parameters:

$$\frac{dz}{dt_0} = -(\mu - \nu) - z + \frac{1}{Q}(1 - z^2)\, r_1(t_0), \tag{1}$$

where $z(t_0) = w^+_{ps}(t_0) - w^-_{ps}(t_0)$, with w^+_{ps} the a posteriori probability of the signal being in the state +a, while w^-_{ps} is the a posteriori probability of the signal being in the state $-a(-1 < z(t_0) < +1)$; $r_1(t_0) = r(t_0)/a = s_1(t_0) + n_1(t_0) = s(t_0) + n(t_0)$ is the signal at the input of the filtering system); $t_0 = t(\alpha + \beta)$ is the dimensionless time; $\mu = \alpha/\alpha + \beta)$ — probability of the signal staying in the state -a (pause); $\nu = \beta/\alpha + \beta)$ — probability of the signal being in the state +a (signal); $Q = N(\alpha + \beta)/a^2$ — generalized noise/signal ratio.

The decision element that concludes the character of the useful signal on the basis of the criterion of the maximal a posteriori probability can be a relay which responds in definite fashion to the reversal of the sign of the function $z(t_0)$, and which forms a temporal sequence of pulses $\tilde{s}_1(t_0)$. A diagram of the optimal receiver is shown in Fig. 1.

For the dimensionless parameters indicated above, we can write the following equation for linear filtering, on the basis of the Kolmogorov-Wiener theory ([5], p. 273):

$$\frac{dz}{dt_0} = -\sqrt{1 + \frac{8\,\mu\nu}{Q}}\, z + \frac{8\,\mu\nu}{Q}\, \frac{r_1(t_0)}{1 + \sqrt{1 + \frac{8\,\mu\nu}{Q}}}. \tag{2}$$

The process $z(t_0)$ at the output of the optimal linear filter is a "smoothed" signal $r(t_0)$, filtered in accordance with the minimum mean-square error criterion. The output of the linear filter is connected to a relay which forms a temporal sequence $\tilde{s}_2(t_0)$. It is quite interesting to compare the filtering efficiencies of the nonlinear and "linear" systems* for different signal/noise ratios, for different signal parameters, and also for a variable relay operation threshold.

---

*The term "linear system" is arbitrarily applied to the aggregate of a linear filter and a relay, which actually is a nonlinear system.

2. PROBABILITY OF FALSE DETECTION AND OMISSION IN THE CASE OF NONLINEAR AND LINEAR FILTERING. APPROXIMATE CALCULATION METHODS

In view of the difficulty in the analytic investigation of the quality of filtering in the general case ($\mu \sim \nu$), we consider the filtering of a sharply asymmetrical signal ($\mu \gg \nu$). If we choose an observation time interval T_0 satisfying the condition $1/\mu \ll T_0 \ll 1/\nu$, then the problem of detecting the signal assumes the same character as the problem of separating a single pulse from noise. The a priori probability density of the signal duration is expressed here by the relation $w^+ = \mu \exp(-\mu\tau)$, In the process of separating the signal from the noise we shall deal with an omission error D, occurring when the system does not respond to a change in the signal polarity, and a false-detection error P_F, when the system indicates a change in polarity of the signal when no such change takes place. It is interesting to estimate these errors.

In fact, the summary mean-squared filtering error, which was chosen as the criterion, is proportional to these errors:

$$\int_0^{T_0} |s(t) - s_0(t)|^2 \, dt = 4a^2 [pT_p D + T'_p P_F],$$

where p — a priori probability of appearance of a pulse; T_p — its average duration; T'_p — average duration of false pulse.

In our case we can assume that the probability distribution of the signal at the output of the filters, ahead of the relay, is stationary in the interval between the positive pulses (during the time of the pause we have $s_1 = -1$), and the average number of false pauses per unit time can be determined as the flux of the probability of the stationary distribution through the boundary $z = z_b$.

Let us find the stationary distribution of the probabilities at the output of a nonlinear and linear filter during the time of the pause. Our system, which is made up of a nonlinear and linear filter, is described by an equation of the type

$$\dot{y} = f(y) + \zeta(t_0). \tag{3}$$

Here $\zeta(t_0)$ is a δ-correlated random process with $\langle \zeta \rangle = 0$ and $\langle \zeta\zeta_\tau \rangle = N_0 \delta(\tau)$. In fact, let us divide both halves of (1) by $1 - z^2$ and let us introduce a new variable y, connected with z by the relation $y = \tanh^{-1} z$. Then the equation of the nonlinear filtering (1) is represented in the form (3), where

$$f(y) = -(\mu - \nu)\,\mathrm{ch}^2 y - \frac{1}{2}\,\mathrm{sh}\, 2y - \frac{1}{Q}; \quad \zeta = \frac{n_1}{Q}; \quad N_0 = \frac{1}{Q}\,(r_1 = -1 + n_1).$$

Using the Fokker-Planck equation we can obtain for the process at the output of the filter $y(t_0)$ the following stationary probability distribution:

$$w_0(y) = C \exp\left\{\frac{2}{N_0}\int f(y)\, dy\right\}, \tag{4}$$

where

$$C = \left[\int \exp\left\{\frac{2}{N_0}\int f(y)\, dy\right\} dy\right]^{-1}.$$

We finally obtain the following expression for the probability density of the values of the signal at the output of the nonlinear filter during the time of the pause

$$w_0^{\text{n}}(z) = \frac{\exp\left\{-\left(1+\frac{\mu-\nu}{2}Q\right)\operatorname{ar\,th}(\mu-\nu)\right\}}{K_{1+\frac{\mu-\nu}{2}}(Q\sqrt{\mu\nu})} \frac{(1-z)^{\frac{\mu-\nu}{2}Q}}{(1+z)^{2+\frac{\mu-\nu}{2}Q}} \times$$
$$\times \exp\{-Q\sqrt{\mu\nu}\operatorname{ch}[2\operatorname{ar\,th} z + \operatorname{ar\,th}(\mu-\nu)]\}. \tag{5}$$

The equation for linear filtering assumes directly the form (3)

$$y \equiv z; \quad f(z) = -Az - B;$$
$$\zeta = Bn_1; \quad N_0 = B^2 Q,$$

where

$$A \equiv \sqrt{1+\frac{8\mu\nu}{Q}}; \quad B = \frac{8\mu\nu}{Q}\frac{1}{1+\sqrt{1+\frac{8\mu\nu}{Q}}}.$$

Using (4), we can readily obtain the stationary probability distribution density for the signal at the output of the linear filter

$$w_0^l(z) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(z-m)^2}{2\sigma^2}\right\}, \tag{6}$$

where

$$m = -\frac{B}{A}; \quad \sigma^2 = \frac{B^2}{2A}Q.$$

We can now proceed to determine the error due to false detection of a positive pulse, occurring during operation of the nonlinear and linear filters, using the approximate solution of the Fokker-Planck equation used in [7, 8]. Let us consider first the case when the threshold of operation of the relay at the output of the filters is zero. During the time of the pause the useful signal acting on the input of the filtering system is $s_1 = -1$. Then the region under consideration is that of negative values of $y(t_0)$, i.e., $-\infty < y(t_0) < 0$ ($y = \tanh^{-1} z$ for nonlinear filtration and $y = z$ for linear filtration). However, there is a finite probability of false detection of a positive pulse, owing to the fact that at a certain instant of time $t = t_1$ the noise trajectory at the output of the filter passes through the boundary $y_b = 0$ into the positive domain. From this point of view, it is interesting to estimate the average time necessary to reach the boundary. For this purpose we must consider all possible trajectories of the representative point in the negative domain, with the exception of the trajectories near the boundary $y(t_0) = y_b$, which certainly do not touch each other, since the speed of motion of the representative part is infinite ($dy(t_0)/dt_0$ has an infinite dispersion). It is, therefore, necessary to put $w_s(y_b) = 0$, where $w_s(y)$ is the somewhat modified stationary probability distribution $w_0(y)$. It can be found as a solution of the Fokker-Planck equation with this boundary condition. The Fokker-Planck equation has the following form:

$$\frac{\partial w(y,t)}{\partial t} = -\frac{\partial G(y,t)}{\partial y},$$

where $G(y, t) = f(y)w(y, t) - \frac{N_0}{2}\frac{\partial w(y,t)}{\partial y}$ has the meaning of a probability flux. As follows from [7, 8], the probability of false detection during the time T_0 is

$$P_F(T_0) = 1 - e^{-\gamma T_0},$$

where

$$\gamma = G_s(y)\big|_{y=y_b} = f(y)w_s(y)\big|_{y=y_b} - \frac{N_0}{2}\frac{dw_s(y)}{dy}\bigg|_{y=y_b},$$

or

$$\gamma = -\frac{N_0}{2}\frac{dw_s(y)}{dy}\bigg|_{y=y_b} \tag{7}$$

is the flux of the probability of stationary distribution through the boundary $y_b = 0$. It is easy to show that γ is the reciprocal of the mean time necessary for the trajectory of the representative point to reach the boundary, or the mean number of transitions of the trajectory $y(t_0)$ through the boundary per unit time; it can serve as an approximate estimate of the number of pulses per unit time.

If the noise is sufficiently small, it can be assumed that the influence of the zero boundary condition on the variation of the probability distribution density is confined to a narrow region near the boundary. We shall, therefore, represent $w_s(y)$ as the product of the stationary probability $w_0(y)$, which satisfies the condition of vanishing at infinity, and a correction factor $w_1(y)$:

$$w_s(y) = w_0(y)\, w_1(y), \tag{8}$$

with $w_1(0) = 0$, and $w_1(y) = 1$ for $y \gg 1/\sqrt{\lambda K}$, where $\lambda = 2/N_0$; $K = -\frac{df(y)}{dy}\Big|_{y=y_b}$. This last inequality imposes a limitation on the noise [7, 8]. Using (7) and (8) we get

$$\frac{dw_s(y)}{dy}\bigg|_{y=0} = w_0(y)\frac{dw_1(y)}{dy}\bigg|_{y=0} = -\frac{2\gamma}{N_0},$$

i.e., after integration,

$$w_1(y)\Big|_{|y| > \frac{1}{\sqrt{\lambda K}}} = -\frac{2\gamma}{N_0}\int_0^y \frac{1}{w_0(y)}\,dy = 1. \tag{9}$$

Using (9), in which we expand the integrand in a Taylor series about $y = 0$, and let it approach $-\infty$, we can readily find the number of false pulses per unit time for the case of nonlinear filtering

$$\gamma_n = \frac{\mu - \nu + \frac{1}{Q}}{K_{1+\frac{\mu-\nu}{2}Q}(Q\sqrt{\mu\nu})}\, e^{-\frac{Q}{2}} \left[\frac{1+(\mu-\nu)}{1-(\mu-\nu)}\right]^{-\frac{1}{2}\left(1+\frac{\mu-\nu}{2}Q\right)} \tag{10}$$

If the operating threshold of the relay b is not equal to zero, we obtain a more complicated expression

$$\gamma_n = \frac{\left[\frac{1+(\mu-\nu)}{1-(\mu-\nu)}\right]^{-\frac{1}{2}\left(1+\frac{\mu-\nu}{2}Q\right)}}{K_{1+\frac{\mu-\nu}{2}Q}(Q\sqrt{\mu\nu})}\left[\left(\frac{1}{Q}+\frac{\mu-\nu}{2}\right)+\frac{\mu-\nu}{2}\,\frac{1+b_1^2}{1-b_1^2}-\frac{b_1}{1-b_1^2}\right]\times$$
$$\times\left(\frac{1+b_1}{1-b_1}\right)^{-\left(1+\frac{\mu-\nu}{2}Q\right)}\exp\left\{-Q\left[\frac{1}{2}\,\frac{1+b_1^2}{1-b_1^2}+(\mu-\nu)\frac{b_1}{1-b_1^2}\right]\right\}, \tag{11}$$

where $b_1 = b/a$. At the same time we have for linear filtering

$$\gamma_l = \frac{A}{B}\sqrt{\frac{A}{\pi Q}}\left(b_1+\frac{B}{A}\right)\exp\left\{-\frac{A}{B^2Q}\left(b_1+\frac{B}{A}\right)^2\right\}. \tag{12}$$

It is quite interesting also to determine the probability of omission of positive pulses $D(\tau)$ as a function of their duration τ for both the nonlinear and linear filterings. Inasmuch as we deal here with a sharply asymmetrical useful signal, we can assume that at the instant $t_0 = 0$, when we assume the signal to jump from -1 to $+1$, a stationary voltage has already been established at the output of the filter.

The equation of the trajectory of the representative point at the output of the nonlinear filter is, in the absence of noise,

$$Z(t_0, z_0) = z_1 + \left\{\left[\frac{1}{2z_1+Q} + \frac{1}{z_0 - z_1}\right] e^{\left(\frac{2z_1}{Q}+1\right) t_0} - \frac{1}{2z_1+Q}\right\}^{-1},$$

where $z_1 = -\frac{Q}{2} + \sqrt{\frac{Q^2}{4} - Q(\mu - \nu) + 1}$ is the voltage at the output of the nonlinear filter in the absence of noise, under the condition where the filter is acted upon for a long time by the signal $s_1 = +1$; $z_0 \equiv z(0) = z_{-1} = \frac{Q}{2} + \sqrt{\frac{Q^2}{4} + Q(\mu - \nu) + 1}$ — voltage at the output of the nonlinear filter during the time of the pause ($s_1 = -1$); z_1 and z_{-1} are particular solutions of the Eq. (1) in the absence of noise. Obviously, the delay t^* in the observation of the positive pulses in the absence of noise is determined from the relation $Z(t^*, z_{-1}) = b_1$.

In the presence of noise, z_0 is a random quantity. We shall assume that the noise is small and manifests itself essentially in the initial distribution; we shall assume also that it does not influence the trajectory of the representative point. Assume that at the instant of time τ the trajectory passes into the positive region of values through the boundary $Z(\tau, z_0) = b_1$, and a positive pulse is observed. From this relation we can readily find the "start" of the trajectory $Z_0(\tau)$. It is obvious that under the assumptions made above, the probability of nonobservation of a pulse with duration τ and shorter is

$$D_n(\tau) = \int_{-1}^{Z_0(\tau)} w_0^{n}(z)\, dz, \tag{13}$$

where $w^n_0(z)$ is the stationary probability of the distribution of noise at the output of a nonlinear filter acted upon by a signal $s_1 = -1$ with noise.

Let us proceed now to linear filtering. Assuming that at the instant of time $t_0 = 0$ the value assumed by s_1 is $+1$, we can readily obtain the following equation for the trajectory of the representative point at the output of a linear filter in the absence of noise:

$$Z(t_0, z_0) = \left(z_0 - \frac{B}{A}\right) e^{-At_0} + \frac{B}{A}.$$

Under the same assumptions as above, we can find the probability of nondetection of pulses of duration τ and shorter with a linear filter

$$D_l(\tau) = \int_{-\infty}^{Z_0(\tau)} w_0^{l}(z)\, dz,$$

where $w^l_0(z)$ is the stationary density of distribution of probabilities when $s_1 = -1$; $Z_0(\tau)$, which is the start of the trajectory, is obviously

$$Z_0(\tau) = b_1 e^{A\tau} - \frac{B}{A}\left(e^{A\tau} - 1\right)$$

and

$$D_l(\tau) = \frac{1}{2} + \frac{1}{2}\Phi\left[\frac{b_1}{B}\sqrt{\frac{2A}{Q}}\, e^{A\tau} + \sqrt{\frac{2}{AQ}}\left(2 - e^{A\tau}\right)\right]. \tag{14}$$

As τ varies from zero to ∞, $D_l(\tau)$ varies from a value close to unity to zero, under the condition $\frac{b_1}{B}\sqrt{\frac{2A}{Q}} < \sqrt{\frac{2}{AQ}}$. This inequality imposes a limitation on the relay operating threshold, namely $b_1 < B/A$. Actually, when $b_1 = B/A$, when there is no influence of noise on the trajectory (which is what we assume), the pulse will be observed at the instant of time $\tau \to \infty$.

The foregoing theoretical estimates (10) and (12) — (14) were numerically evaluated for several specific cases ($Q = 0.22$; $Q = 0.44$; $\mu = 0.9$, and $\nu = 0.1$), which were investigated experimentally. In [6] are given and discussed the results of the experiment and the theoretical estimates obtained for these results. The theoretical estimates were in quite satisfactory agreement with experiment.

As was shown above, γ is the number of sets of transitions of the noise trajectory at the filter output through the boundary per unit time. Inasmuch as the relay has a certain "dead" time and does not respond to each instantaneous transition of the trajectory through the boundary, it can be assumed that γ yields an approximate estimate of the number of false pulses per unit time. According to the experimental data, for linear filtering, γ_{exp} exceeds γ by a factor of 1.2 when $Q = 0.22$ and by a factor of 1.3 when $Q = 0.44$, with $b_1 = 0$. On the other hand, γ obtained for nonlinear filtering differs extremely little from γ_{exp}.

The theoretical estimates obtained for the probability of omission of a signal, assuming that the noise does not influence the system at the instant when the leading front of the pulse is in operation, were compared with the experimental data. It was found that the effect of noise on the trajectory during the time of signal sign reversal leads to a considerable reduction in the probability of omission of a signal for small values of τ. At sufficiently large τ, the theoretical estimates are in good agreement with the experimental results [6].

3. COMPARISON OF OPERATION OF NONLINEAR AND LINEAR FILTERING SYSTEMS IN THE CASE OF SMALL NOISE ($Q \ll 1$)

Let us consider the problem of finding the characteristics of optimal receivers for $Q \ll 1$ under the additional assumption that $\nu/Q \ll 1$. Obviously, this condition is satisfied almost always, even for very small Q, in the case when a single pulse is filtered ($\nu \to 0$).

The equation for optimal nonlinear filtering (1) can be represented in the form

$$\frac{dv}{dt_0} + v\left(1 - \frac{2r_1(t_0)}{Q}\right) = \frac{4\nu}{Q} r_1(t_0),$$

where $\nu = 1 + z - 2\nu$ (it is assumed that $Q \ll 1$; $\nu \ll 2$). The equation for optimal nonlinear filtering for the case under consideration $\nu/Q \ll 1$ has the form

$$\frac{dz}{dt_0} + z = \frac{4\nu}{Q} r_1(t_0).$$

From the last two equations we see that filtering systems behave in principle in different fashion at small values of Q, and the smaller Q the greater this difference.

It will prove convenient in what follows to consider the equation of nonlinear filtering in the form

$$\dot{u} = -u + 2\nu + \frac{2}{Q} u r_1 \qquad (Q \ll 1), \tag{15}$$

where $u = 1 + z$ (it is assumed that $z < 0$, $u \ll 2$).

During the time of the pause the voltage at the output of the nonlinear filter has the following distribution, which is derived from (15):

$$w_0(u) = \frac{(\nu Q)^{\frac{Q}{2}+1}}{\Gamma\left(\frac{Q}{2}+1\right)} u^{-\left(\frac{Q}{2}+2\right)} e^{-\frac{\nu Q}{u}}. \tag{16}$$

Using the method considered above, we readily obtain the following theoretical estimate for the number of false pulses per unit time:

$$\gamma_n = \frac{2\nu}{1+b_1}. \tag{17}$$

Let us find the probability of omission of a signal for nonlinear filtering. The equation of the trajectory at the output of the filter during operation of the signal has the following form:

$$u(t_0) = -\frac{2\nu}{\frac{2}{Q}-1} + \left[u(0) + \frac{2\nu}{\frac{2}{Q}-1} \right] e^{\left(\frac{2}{Q}-1\right) t_0}.$$

The statistical spread u(0) can be disregarded, for the width of the probability distribution curve is on the order of νQ, and when $\nu Q \ll 1 + b_1$ we can assume that u(0) is a constant quantity, $u(0) = \nu Q$. The inequality $b_1 \gg -1 + \nu Q$ imposes a condition on the quantity b_1. At the instant of time t_1, determined from the condition $u(t_1) = 1 + b_1$, the pulse is observed. The probability of omission (the probability of a useful-signal pulse has a shorter duration than t_1) is expressed in the form $D_n = 1 - e^{-t_1}$, or ultimately

$$D_n = 1 - (2\nu Q)^{\frac{Q}{2}} (1 + b_1)^{-\frac{Q}{2}}. \tag{18}$$

Eliminating b_1 from (17) and (18) we obtain the following expression for the interdependence of the nonlinear-filtering errors

$$\gamma_n = \frac{1}{Q} (1 - D_n)^{2/Q}. \tag{19}$$

At the same time, we can obtain from (12) the following theoretical estimate for the number of false pulses per unit time in the case of linear filtering, taking the conditions $\mu \gg \nu$, $Q \ll 1$, $\nu/Q \ll 1$ into account:

$$\gamma_l = \frac{1}{\sqrt{\pi}} \frac{b_0 \sqrt{Q}}{4\nu} e^{-\frac{b_0^2 Q}{16\nu^2}}, \tag{20}$$

where $b_0 = b_1 + B/A$. In our case $b_0 = b_1 + 4\nu/Q$. The equation of the trajectory at the output of the linear filter, during a positive pulse, has the following form:

$$\frac{dz}{dt_0} + z = \frac{4\nu}{Q} r_1(t_0).$$

Let $b_0 \gg \sigma$ (low noise). We can then assume that all the trajectories start from $z(0) = 4\nu/Q$. As a result

$$D_l = \frac{b_0 Q}{8\nu}. \tag{21}$$

The characteristic of detecting a signal by a linear system can be readily obtained in the form

$$\gamma_l = \frac{2}{\sqrt{\pi}} \frac{D_l}{\sqrt{Q}} e^{-\frac{4D_l^2}{Q}}. \tag{22}$$

It must be noted that the condition $\nu/Q \ll 1$ is used only in the derivation of formulas (20)-(22). In those cases when $\nu/Q \sim 1$ or $\nu/Q > 1$, the characteristic of signal detection by means of a linear system is described by the relation

$$\gamma_l = \frac{2}{\sqrt{\pi}} \frac{A\sqrt{A}}{\sqrt{Q}} D_l e^{-\frac{4AD_l^2}{Q}}. \tag{23}$$

Let us find the limit of applicability of formula (20), and hence of formulas (22) and (23). As already indicated above, we have dealt in the derivation of formula (9) with a somewhat

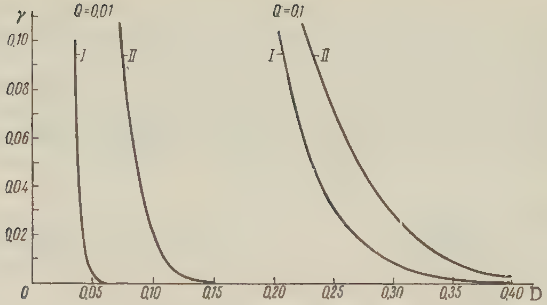


Fig. 2. Detection characteristics of a nonlinear filtering system (I) and of a linear one (II).

modified probability density distribution function. The correction factor is equal to unity outside the narrow region along the boundary, the width of which is $1/\sqrt{\lambda K}$, where for the case of linear filtering $\lambda = 2/B^2Q$ and $K = 1$, i.e., $1/\sqrt{\lambda K} = \sigma$.

Formula (20) is obviously valid for $b_1 - (-B/A) > 2\sigma$ or $b_0 > 2\sigma$, i.e., we have obtained a condition for the threshold of the relay connected to the output of the linear filter, a condition equivalent in its sense to the condition for the operating threshold of the relay in the nonlinear system, as obtained above. The latter inequality corresponds to the conditions $\gamma_l < 0.108$ and $D_l > \sqrt{Q/2}$. The range of applicability of (17) and (19) is much broader.

Fig. 2 shows the detection characteristics of the nonlinear and linear systems for $Q = 0.1$ and $Q = 0.01$; $\nu/Q \ll 1$. We see from these characteristics that the nonlinear system has an undisputed advantage over the linear one in the case of filtering of a single pulsed signal of random duration, when $Q \ll 1$. The advantage of the nonlinear filtering is all the greater, the smaller Q. Using these plots we can find the connection between the probabilities of omission and of false detection of a single pulse during an observation time T_0. The probability of the latter error, P_F is expressed in terms of γ by the formula $P_F = 1 - e^{-\gamma T_0}$. The time of observation T_0 is subject to the condition $1/\mu \ll T_0 \ll 1/\nu$.

CONCLUSIONS

1. In the case of filtering of a sharply asymmetrical rectangular random signal out of a white noise, the foregoing optimal nonlinear and linear filters are characterized by false-detection and omission errors, which can be determined from formulas (10), (12) — (14). Theoretical estimates of the errors arising during the filtering are in satisfactory agreement with the experimental results of [6].

2. On the basis of the formulas obtained for the estimates of the filtering errors, we can plot the characteristics of detection for various signal-to-noise ratios, and also for different signal parameters.

3. If the observation time T_0 satisfies the condition $1/\mu \ll T_0 \ll 1/\nu$ (in the case of filtering of a single random pulse with $Q \ll 1$, the nonlinear filter is better than the linear one, the difference between the two increasing with decreasing Q.

REFERENCES

1. A.N. Kolmogorov. Interpolation and extrapolation of stationary and random sequences. Izv. AN SSSR (News, Academy of Sciences USSR), Mathematics Series, 1941, 5, 1, 3.
2. N. Wiener. The Extrapolation, Interpolation and Smoothing of Stationare Times Series. J. Wiley, N.Y., 1949.
3. R.L. Stratonovich. On the theory of optimal nonlinear filtering of random functions. Teoriya Veroyat. i Primen. 1949, 4, 2, 289.
4. R.L. Stratonovich. Conditional Markov Process, ibid, 1960, 5, 2, 172.
5. V.V. Solodovnikov. Introduction to the Statistical Dynamics of Automatic Control Systems, GITTL, 1952.
6. N.K. Kul'man, P.S. Landa. Use of electronic models to investigate certain optimal devices for filtering pulse signals of random duration, Radiotekhnika i elektronika, 1961, 6, 4, 506.

7. R.L. Stratonovich, P.S. Landa. Effects of noise on the generator with hard excitation. Izvestiya Vuzov, Ministry of Higher Education USSR — Radiophysics, 1959, 2, 1, 37.
8. P.S. Landa. On the stability of systems with servocontrol in the presence of random signals, Automation and Remote Control, 1960, 21, 1, 36.
9. R.L. Stratonovich. Use of the theory of Markov processes for optimum filtration of signals, Radiotekhnika i elektronika, 1960, 5, 11, 1751.

Physics Faculty
M.V. Lomonosov
Moscow State University

Received by editor 26 October 1960

EFFECT OF SIGNAL FLUCTUATIONS ON A RECEIVER WITH AGC

V.V. Shirokov

We consider the effect of a signal of fluctuating sinusoidally modulated amplitude on a receiver with AGC. The passage of such a signal through the receiver is analyzed for the case when additional noise components are present near the sinusoidal-modulation frequency. The statistical characteristics of the random process at the output of the receiver are determined and considerations are advanced concerning the choice of the time delay produced by the AGC filter.

INTRODUCTION

In many practical cases it is necessary to determine the effect of a sinusoidally modulated signal with fluctuating amplitude on a receiver with automatic gain control (AGC). The AGC system, acting on the envelope of the signal, should eliminate the fluctuations as much as possible and to change the least the sinusoidal component, for this component contains the useful information. It is, therefore, of interest to ascertain the statistical characteristics of the output voltage of the receiver and to discuss a rational choice of time delay of the AGC filter.

In some cases additional noise components appear near the sinusoidal modulation frequency. This situation obtains, for example, in radar [2, 3], where these additional noise components are called "angle noise." An estimate of the effect of the AGC on angle noise is found in [3]. Rigorous analysis of the effect of signal fluctuations on real AGC systems, however, is not given in this paper. Recommendations are made on the choice of the AGC system parameters, starting only from the requirement that the effect of angle noise be reduced as much as possible, but no account is taken of the fashion by which real AGC systems eliminate the amplitude fluctuations of the signal.

In the present paper we analyze the effect of a fluctuating sinusoidally modulated signal on a receiver with AGC. The problem is solved by successive approximation both in the presence of angle noise and in its absence. The mathematical expectation and the spectral density of the random process at the output of the receiver are determined, and recommendations are made on the choice of the AGC system parameters.

1. SOLUTION OF THE INTEGRAL EQUATION DESCRIBING THE AGC

Fig. 1 shows the equivalent circuit of an AGC circuit with delay. The signal from the receiver output $U_{out}(t)$ is fed to the input of the feedback circuit, the output of which develops a regulation voltage $E_r(t)$ which varies the gain of the receiver in such a way as to ensure small deviations of the signal at the output in spite of large changes in the input signal.

In view of the large delay time of the AGC, the regulated receiver itself can be regarded as inertialess and the input disturbance $U_{in}(t)$ is assumed to be the envelope of the input signal even when the latter is of the pulse type. The feedback circuit is assumed linear with respect to the envelope, with a transfer function $k_1 H(\omega)$. Approximating the regulation characteristic of the amplifier within the limits of the working range, corresponding to the steady state of the AGC at a given input-signal level, by means of a straight line we obtain

$$U_{out}(t) = U_{in}(t)\,[k_0 - bE_r(t)]. \qquad (1)$$

We represent the input signal in the form of a Fourier integral

$$U_{in}(t) = A + \frac{1}{2\pi}\int_{-\infty}^{\infty} a(\omega)\, e^{j\omega t}\, d\omega, \qquad (2)$$

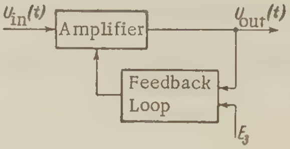


Fig. 1.

where $A = \overline{\overline{U_{in}(t)}}$ is the mathematical expectation of the input signal, averaged over the time.

The output voltage and the control voltage will be written in the form

$$U_{out}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} C(\omega)\, e^{j\omega t}\, d\omega; \quad E_r(t) = \frac{k_1}{2\pi}\int_{-\infty}^{\infty} H(\omega)\, C(\omega)\, e^{j\omega t}\, d\omega - k_1 E_d, \qquad (3)$$

where E_d is the delay voltage.

From the mathematical point of view, the representation of a random process in the form of an ordinary integral is not rigorous, but it is acceptable, since we shall be interested in the course of the solution in the mathematical expectation and in the spectral density of the process, which are connected with the random complex spectrum $C(\omega)$ by the simplest relations (11) and (13).

From (1), (2), and (3) we can readily obtain an integral equation relating the steady-state spectra of the voltages at the output of the receiver and its input:

$$C(\omega) = \frac{k_0 + bk_1E_d}{1 + Abk_1H(\omega)}\,[2\pi A\delta(\omega) + a(\omega)] - \frac{bk_1}{2\pi}\int_{-\infty}^{\infty} \frac{C(s)\, a(\omega - s)\, H(s)\, ds}{1 + Abk_1H(\omega)}, \qquad (4)$$

where $\delta(\omega)$ is the delta-function.

We present the solution of this equation, obtained in [4] by successive approximation assuming the ratios $\sigma_a^2/A^2 \ll 1$ and $m^2 \ll 1$ to be small (σ_a^2 is the dispersion of the fluctuation and m is the depth of the sinusoidal modulation of the input signal):

$$C(\omega) = C_0(\omega) + C_1(\omega) + C_2(\omega) + C_3(\omega) + \ldots,$$

where

$$\begin{gathered} C_0(\omega) = k_{av} A\, 2\pi\delta(\omega); \quad C_1(\omega) = H_{eq}(\omega)\, a(\omega); \\ C_2(\omega) = -H_{eq}(\omega)\,\frac{b}{2\pi}\int_{-\infty}^{\infty} a(s)\, a(\omega - s)\, H_r(s)\, ds; \\ C_3(\omega) = H_{eq}(\omega)\left(\frac{b}{2\pi}\right)^2 \iint_{-\infty}^{\infty} a(s')\, a(s - s')\, a(\omega - s)\, H_r(s)\, H_r(s')\, ds\, ds' \end{gathered} \qquad (5)$$

Here

$$k_{av} = \frac{k_0 + bk_1E_d}{1 + Abk_1}; \qquad (6)$$

$$H_{\mathrm{eq}}(\omega) = \frac{k_{\mathrm{av}}}{1 + Abk_1 H(\omega)}; \tag{7}$$

$$H_{\mathrm{r}}(\omega) = \frac{k_1 H(\omega)}{1 + Abk_1 H(\omega)}. \tag{8}$$

This solution makes it possible to obtain the statistical characteristics of the random process at the output of the receiver in the cases of interest to us.

2. EFFECT OF SIGNAL FLUCTUATIONS ON THE AGC SYSTEM

The signal at the input of the receiver, and the absence of angle noise, can be represented in the form

$$\begin{aligned} U_{\mathrm{in}}(t) &= A(1 + m\cos\Omega t)\,[1 + \xi(t)] = \\ &= A + Am\cos\Omega t + A\xi(t) + A\xi(t)\,m\cos\Omega t, \end{aligned} \tag{9}$$

where $A\xi(t)$ is a stationary random process and Ω is the frequency of the sinusoidal modulation of the signal.

As can be seen from (9), the signal at the input is a nonstationary random process, the mathematical expectation and dispersion of which are periodic functions of the time. Thus, the problem reduces to an investigation of the effect of the nonstationary random process on a nonlinear system. To find an exact solution is quite difficult in this case. We can, therefore, use successive approximations of (5), obtained for the voltage spectrum at the output, where in lieu of the function $a(\omega)$, which is the spectrum of the AC component of the input signal, we substitute in accordance with (9)

$$\begin{aligned} a(\omega) = a_1(\omega) &+ Am\pi\,[\delta(\omega + \Omega) + \delta(\omega - \Omega)] + \\ &+ \frac{m}{2}\,[a_1(\omega + \Omega) + a_1(\omega - \Omega)], \end{aligned} \tag{10}$$

where $a_1(\omega)$ is the spectrum of the process $A\xi(t)$.

The solution obtained in this manner for $C(\omega)$ enables us to determine the characteristics of the random process of interest to us, at the output of the receiver. Actually, the mathematical expectation of the voltage at the output can be found from the relation

$$\overline{U_{\mathrm{out}}(t)} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{C(\omega)}\,e^{j\omega t}\,d\omega. \tag{11}$$

Substituting the expression for $C(\omega)$, we find that the mathematical expectation of the output signal, accurate to terms of third order, will have the form

$$\begin{aligned} \overline{U_{\mathrm{out}}(t)} = k_{\mathrm{av}}A\Big\{1 &- \frac{m^2Ab}{2(1 + Abk_1)}\,\mathrm{Re}\,[H_{\mathrm{r}}(\Omega)] - \\ - \frac{b}{A(1 + Abk_1)\,2\pi}\int_{-\infty}^{\infty} H_{\mathrm{r}}(s)\,g(s)\,ds &+ \frac{m}{k_{\mathrm{av}}}\,\mathrm{Re}\,[H_{\mathrm{eq}}(\Omega)\,e^{j\Omega t}] + \\ &+ \frac{m^2Ab}{2k_{\mathrm{av}}}\,\mathrm{Re}\,[H_{\mathrm{eq}}(2\Omega)\,H_{\mathrm{r}}(\Omega)\,e^{2j\Omega t}]\Big\}. \end{aligned} \tag{12}$$

In the derivation of this expression we used the relation, which is obvious for a stationary random process,

$$\overline{a_1(\omega)\,a_1(\omega')} = g(\omega)\,2\pi\delta(\omega + \omega'),$$

where $g(\omega)$ is the spectral density of the fluctuations $A\xi(t)$.

The random process at the output will be nonstationary. In view of the fact that at the output of the receiver one connects usually circuits with large time delay, circuits which average the random process with respect to time, a sufficient description of the output voltage will be provided by the spectral density, corresponding to the time-averaged correlation function. Then, as was done in [4], the spectral density $g_{\text{out}}(\omega)$ can be found from the relation

$$\overline{\overline{[C(\omega)-\overline{C(\omega)}]\,[C^*(\omega')-\overline{C^*(\omega')}]}} = 2\pi g_{\text{out}}(\omega)\,\delta(\omega-\omega'). \tag{13}$$

The wavy line denotes that we choose in the mathematical expectation the terms corresponding to the time-averaged correlation function.

Substituting in (13) the solution for $C(\omega)$, we obtain the first approximation for $g_{\text{out}}(\omega)$ in the form

$$g_{\text{out}1}(\omega) = |H_{\text{eq}}(\omega)|^2 g(\omega). \tag{14}$$

Thus, in the first approximation the AGC system acts on the signal fluctuations as a linear system with a transfer function determined by formula (7). In analogy we can obtain the correction to the first approximation. We shall not give the complete expression for the value of this correction, since it is too cumbersome, and we shall use the fact that the outputs of these receivers usually are connected to narrowband filters, tuned to the frequency Ω. We shall, therefore, be interested only in the fluctuation components near this frequency. Assuming that the bandwidth of the fluctuation spectrum at the output is narrower than the frequency range $0-\Omega$, we find that near the frequency Ω the spectral density of the output voltage will have the form

$$\Delta g_{\text{out}}(\omega) \simeq |H_{\text{eq}}(\omega)|^2 g(\omega-\Omega)\frac{m^2}{2}\{A^2b^2[|H_{\text{r}}(\omega-\Omega)|^2+|H_{\text{r}}(\Omega)|^2]+$$
$$+2A^2b^2\operatorname{Re}[H_{\text{r}}(\omega-\Omega)H_{\text{r}}(\Omega)]-2Ab\operatorname{Re}[H_{\text{r}}(\omega-\Omega)+H_{\text{r}}(\Omega)]+1\}. \tag{15}$$

Formula (15) enables us to calculate the spectral density of the voltage near the frequency Ω for all ordinary parameters of the AGC system. In the case when these parameters are chosen so as to obtain a minimum distortion of the sinusoidal envelope of the signal, we should, as follows from (12), (7) and (8), satisfy the conditions

$$\begin{aligned}&|H_{\text{eq}}(\Omega)|^2 \simeq k_{\text{av}}^2;\\&|H_{\text{r}}(\Omega)| \simeq 0.\end{aligned} \tag{16}$$

If these conditions are satisfied, formula (15) simplifies considerably:

$$\Delta g_{\text{out}}(\omega) \simeq \frac{m^2}{2} g(\omega-\Omega)\,|H_{\text{eq}}(\omega-\Omega)|^2. \tag{17}$$

It should be noted that formulas (15) and (17) are obtained under the assumption that the spectrum of the fluctuations at the input, $A\xi(t)$, has such a narrow bandwidth that the components of these fluctuations near the frequency Ω are practically equal to zero. But this assumption may not be satisfied, since for small values of m the components $\Delta g_{\text{out}}(\omega)$ which we have calculated may prove to be comparable with or less than the components of the fluctuations $A\xi(t)$ which we obtain near the frequency Ω. These components, however, can be taken into account by using the first-approximation formula.

If conditions (16) are satisfied, then part of the spectral density, due to presence of these components near the frequency Ω, can be regarded as approximately constant and equal to

$$\Delta g_n(\omega) \simeq 2k_{\text{av}}^2 g(\Omega). \tag{18}$$

This means that AGC does not affect at all the components of the spectrum of the fluctuations of the input signal $A\xi(t)$ near the frequency Ω. It can be shown that this estimate of these components is sufficient.

Taking the foregoing into account, we must bear in mind in these formulas and all that follow, that the complete expression for the components near the frequency Ω will be

$$\Delta g(\omega) = \Delta g_{\text{out}}(\omega) + 2k_{\text{av}}^2 g(\Omega). \tag{19}$$

From the relations (14), (15), or (17) it follows that the better the AGC system eliminates the fluctuation components near $\omega = 0$, the better the elimination of the components near $\omega = \Omega$, which occur in the spectrum of the input signal due to the sinusoidal modulation of the fluctuating signal. From this point of view, it is advantageous to reduce the time delay of the AGC system to a limit determined by the permissible distortions of the sinusoidal envelope of the signal. However, the entire preceding analysis was based on the assumption that there are no angle noises, whereas the distance to the target is decreased these begin to play an ever more important role. The characteristics of the fluctuations at the output in the presence of angle noise will be derived in the next section.

3. EFFECT OF SIGNAL FLUCTUATIONS ON THE AGC SYSTEM IN THE PRESENCE OF ANGLE NOISE

The signal at the input of the receiver, in the presence of angle noise, can be represented in the form [3]

$$\begin{aligned} U_{\text{in}}(t) &= A\,[1+\xi(t)]\,[1+m\cos\Omega t] + h(t)\cos\Omega t = \\ &= A + Am\cos\Omega t + A\xi(t) + Am\xi(t)\cos\Omega t + h(t)\cos\Omega t, \end{aligned} \tag{20}$$

where h(t) is a stationary random process with zero expectation value. The component h(t) cos Ωt appears because of the presence of angle noise. We note that the random processes h(t) and ξ(t) are independent.

An analysis of the effect of a signal of this type on the AGC system can be quite analogous to the preceding one. In this case the spectrum of the AC component of the input signal will have the form

$$\begin{aligned} a(\omega) &= Am\pi\,[\delta(\omega-\Omega) + \delta(\omega+\Omega)] + a_1(\omega) + \\ &+ \frac{m}{2}\,[a_1(\omega+\Omega) + a_1(\omega-\Omega)] + \frac{1}{2}\,[b(\omega+\Omega) + b(\omega-\Omega)], \end{aligned} \tag{21}$$

where b(ω) is the spectrum of h(t).

Substituting (21) in (5), we obtain a solution for C(ω), and by using this solution we determine the characteristics of the random process at the output in the case under consideration. It is assumed here that $\sigma_b^2/A^2 \ll 1$, where σ_b^2 is the dispersion of h(t).

The mathematical expectation of the output voltage will be described, accurate to terms of order m^2 and σ_a^2 inclusive, by formula (12), while the first approximation for the spectral density is given by (14). Without citing the complete expression for the spectral density of the signal at the output in the second approximation, we are interested as before only in the fluctuating components at the output near the frequency Ω, for the case of a narrow-band fluctuation spectrum at the output near $\omega = 0$, and in the case of AGC system parameters of the sinusoidal envelope of the signal (16). Then the spectral density of these components will be given by the formula

$$\begin{aligned} \Delta g_{\text{out}}(\omega) &\simeq \frac{m^2}{2}\, g(\omega-\Omega)\, |H_{\text{eq}}(\omega-\Omega)|^2 + \\ &+ \frac{k_{\text{av}}^2}{2}\left[g_b(\omega-\Omega) + \frac{b^2}{2\pi}\int_{-\infty}^{\infty} g(s)\, g_b(\omega-s-\Omega)\, |H_{\text{r}}(s)|^2\, ds\right], \end{aligned} \tag{22}$$

where $g_b(\omega - \Omega)$ is the spectral density of the angle noise at the input.

The first term in this formula represents the spectral density of the fluctuations at the output in the absence of angle noise. The second term represents the spectral density of the angle noise, which would obtain at the output were the AGC not to eliminate the fluctuations at all. We are particularly interested in the last term, which is the convolution of the spectral densities of the control voltage and of the amplitude fluctuations of the input signal. It appears physically because of multiplication of the random components of the control voltage with the angle noise, a multiplication that takes place in the AGC receiver.

Analysis shows that the better the AGC system eliminates the fluctuation components near the frequency $\omega = 0$, the greater the intensity of the additional fluctuation components near $\omega = \Omega$, arising as a result of the angle noise.

Since we may not be interested in the eliminated components near $\omega = 0$, owing to the fact that they do not enter into the pass band of the output filter, it is expedient to increase the inertia of the AGC so as to reduce the effect of the angle noise. It must be borne in mind, however, that this will bring about an increase in the intensity of the components that occur near the frequency $\omega = \Omega$ also in the absence of angle noise (the first term of formula (22)). Therefore, the question of the choice of the inertia of the AGC must be solved by compromise, taking into account the passband of the filter at the output of the receiver, as will be done below using as an example an AGC system with first-order filter.

4. EFFECT OF SIGNAL FLUCTUATIONS ON AN AGC SYSTEM WITH FIRST-ORDER FILTER

Let us consider an AGC system, the feedback loop of which is described by transfer function

$$k_1 H(\omega) = \frac{k_1}{1 + j\omega T}, \tag{23}$$

where T is the time constant of the AGC filter. Let the amplitude fluctuations of the signal (in the absence of angle noise) have the spectral density

$$g(\omega) = 2\alpha\sigma_a^2/(\alpha^2 + \omega^2), \tag{24}$$

which corresponds to a correlation function of the form

$$R(\tau) = \sigma_a^2 e^{-\alpha|\tau|}, \tag{25}$$

where $1/\alpha = T_c$ is the time of fluctuation correlation. Taking (21) into account, we can represent (7) and (8) in the form

$$H_{\text{eq}}(\omega) = k_{\text{av}} \frac{1 + j\omega T}{(1 + Abk_1)(1 + j\omega T_{\text{eq}})}; \quad H_{\text{r}}(\omega) = \frac{k_1}{(1 + Abk_1)(1 + j\omega T_{\text{eq}})}, \tag{26}$$

where $T_{\text{eq}} = T/(1 + Abk_1)$ is the equivalent time constant of the AGC system. Then the mathematical expectation of the process at the output will have the form

$$\overline{U_{\text{out}}(t)} = k_{\text{av}} A \left\{ 1 - \frac{m^2}{2Abk_1(1 + \Omega^2 T_{\text{eq}}^2)} - \frac{\sigma_a^2}{A^3 bk_1(1 + \alpha T_{\text{eq}})} + \right.$$
$$\left. + \frac{m\sqrt{1 + \Omega^2 T^2}}{(1 + Abk_1)\sqrt{1 + \Omega^2 T_{\text{eq}}^2}} \cos(\Omega t - \varphi) \right\}. \tag{27}$$

We have omitted here the terms representing the second harmonic of the regular envelope of the input signal, and the phase shift φ is defined by the formula

$$\varphi = \operatorname{arc\,tg} \frac{\Omega T}{1 + \Omega^2 T^2}. \tag{28}$$

Recognizing that usually $Abk_1 \gg 1$, we can write, with accuracy sufficient for practical purposes,

$$U_{\text{out}}(t) \simeq k_{\text{av}} A \left\{ 1 + \frac{m\sqrt{1 + \Omega^2 T^2}}{(1 + Abk_1)\sqrt{1 + \Omega^2 T_{\text{eq}}^2}} \cos(\Omega t - \varphi) \right\}. \tag{29}$$

This expression coincides with the formula for the output signal, obtained in [1] for the case where there are no fluctuations. This means that if the idealizations introduced here are fulfilled, the fluctuations hardly affect the mathematical expectation of the output signal.

Using (14), we obtain an expression for the components of the fluctuations at the output near $\omega = 0$:

$$g_{\mathrm{out}\,1}(\omega) = \frac{2\sigma_a^2\alpha}{\alpha^2+\omega^2}\,\frac{k_{\mathrm{av}}^2(1+\omega^2T^2)}{(1+Abk_1)^2\,(1+\omega^2T_{\mathrm{eq}}^2)}\,. \tag{30}$$

If the parameters of the AGC system are chosen from condition (16), which in our case reduces to satisfaction of the inequality

$$\Omega T_{\mathrm{eq}} \gg 1, \tag{31}$$

then the total expression for the components of the output fluctuations near the frequency $\omega = \Omega$ are determined by using (14) and (19):

$$\Delta g(\omega) \simeq \frac{m^2\sigma_a^2\alpha}{[\alpha^2+(\omega-\Omega)^2]}\,\frac{k_{\mathrm{av}}^2[1+(\omega-\Omega)^2T^2]}{(1+Abk_1)^2\,[1+(\omega-\Omega)^2T_{\mathrm{eq}}^2]} + \frac{4k_{\mathrm{av}}^2\sigma_a^2}{\alpha^2+\Omega^2}\,. \tag{32}$$

For the more general case of arbitrary T_{eq}, formula (15) must be used. The form of the spectral density of the fluctuations at the output is shown in Fig. 2. From the relations obtained it follows that in the absence of angle noise the smaller the equivalent time constant T_{eq} of the AGC, the better are the fluctuations eliminated both near $\omega = 0$ and near $\omega = \Omega$. The reduction in T_{eq} is limited here by the requirement that the distortion of the sinusoidal envelope of the signal be minimal.

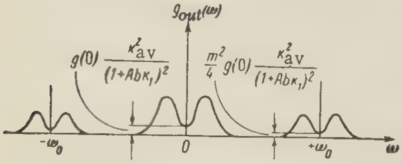


Fig. 2.

In order to estimate the dependence of the intensity of the fluctuations at the output on the parameters of AGC system in the presence of angle noise, let us calculate the dispersion of the voltage at the output of the receiver. Assume that the output of the receiver is connected to a narrowband filter tuned to the frequency Ω, with a bandwidth $\Delta\omega_f = 2\pi\,\Delta f_f$. We specify the spectral density of the angle noise in the form:

$$g_b(\omega-\Omega) = \frac{2\sigma_b^2\alpha}{[\alpha^2+(\omega-\Omega)^2]}\,. \tag{33}$$

To simplify the analysis we assume the correlation time of the amplitude fluctuations to be equal to the correlation times of the angle noise. Recognizing that in practical circuits the bandwidth of the output filter is usually much narrower than the bandwidth of the fluctuation spectrum, we obtain, by using the above-derived relations for $\Delta g(\omega)$, the following expression for the dispersion of the process at the output of the filter

$$\sigma_{\mathrm{out}}^2 \simeq m^2\,k_{\mathrm{av}}^2\sigma_a^2T_c\,\Delta f_f\left\{1-\left[1-\frac{1}{(1+A\,b\,k_1)^2}\right]\frac{2\,\mathrm{arc\,tg}\,\frac{\Delta\omega_f T_{\mathrm{eq}}}{2}}{\Delta\omega_f T_{\mathrm{eq}}}\right\}+$$

$$+k_{\mathrm{av}}^2\,\sigma_b^2\,T_c\,\Delta f_f\left\{1+\frac{\sigma_a^2}{A^2}\,\frac{1+2\alpha T_{\mathrm{eq}}}{2\,(1+\alpha T_{\mathrm{eq}})^2}\right\}+\frac{4k_{\mathrm{av}}^2\sigma_a^2\,\Delta f_f}{\alpha^2+\Omega^2}\,. \tag{34}$$

The first term in this formula represents the dispersion of the output voltage in the absence of angle noise, the second term represents the dispersion component due to angle noise, and the third term is equal to the dispersion of the output voltage resulting from the direct passage of the components of fluctuations of the signal $A\xi$ (t) through the filter (the spectral density of these components at the output of the filter is given by formula (18)).

An analysis of the obtained expression shows that to reduce the influence of angle noise it is necessary to increase the equivalent time constant T_{eq} of the AGC. However, if we choose T_{eq} so large that the condition $\Delta\omega_f T_{\mathrm{eq}} \geq 1$ is satisfied, then the dispersion component proportional to m^2, which exists also in the absence of angle noise, increases rapidly. In addition, the choice of equivalent time constant for the AGC is limited from above by the permissible transient time in the AGC receiver, namely $t_s \simeq 2.3\,T_{\mathrm{eq}}$. Thus, the following conclusions can be drawn regarding the choice of T_{eq}:

1. In the case when there is no angle noise, it is advantageous to reduce T_{eq} to a limit determined by the permissible distortion of the regular envelope of the signal.

2. To reduce fluctuations in the errors due to the angle noise, it is necessary to choose T_{eq} so as to make the ratio T_{eq}/T_c as large as possible.

3. It is undesirable to increase T_{eq} to values for which the condition $\Delta\omega_f T_{eq} \geq 1$ begins to be satisfied, since this increases the fluctuation errors proportional to m^2. In addition, the increase in T_{eq} is limited by the required transient time in the system.

4. In a quantitative estimate of the intensity of the fluctuations at the output, it is necessary to take into account the components of the fluctuations of the signal $A\xi(t)$, components which fall directly in the passband of the output signal, for they can play an appreciable role at small values of m. The AGC does not eliminate these fluctuation components if the parameters of the system are chosen to satisfy the requirement of minimum distortion of the regular signal envelope.

The question of the final choice of T_{eq} must be solved in each specific case by a compromise, depending on the requirements imposed on the receiving apparatus, taking into account the considerations given above.

CONCLUSION

We have considered the effect of a sinusoidally modulated fluctuating signal on a receiver with AGC, having a filter of arbitrary order. We analyzed the case when angle noise is present in addition to the amplitude fluctuations. By way of an example, we consider an AGC system with first-order filter.

REFERENCES

1. G.P. Tartakovskiy, Dynamics of AGC systems, GEI, 1957.
2. S.H. Dunn, D.D. Howard, A.M.King Effect of fluctuations of an echo signal on the operation of a radar target tracking station, Proc. IRE, 47, 855 (1959).
3. R.H. DeLano, I. Pfeffer. Effect of AGC on radar tracking noise, Proc. IRE, 1956, 44, 6, 801.
4. V.V. Shirokov, V.G. Repin. Effect of noise on an AGC system. Radiotekhnika, 1959, 14, 4, 65.

Received by editor 30 June 1960

ON THE DETECTION OF FREQUENCY MODULATED OSCILLATIONS

V.K. Turkin, G.A. Levin

We derive asymptotic formulas which greatly facilitate the calculation of the functions T and U.

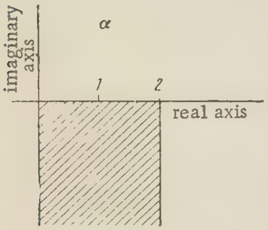
In [1] we introduced new types of special functions, which in our opinion are of interest for the theory of detection of frequency-modulated oscillations. These functions are defined by the formulas

$$T_{\mu}^{(k)}(m, \alpha) = \sum_{n=-\infty}^{\infty} \frac{J_n(m)\, J_{n-\mu}(m)}{(n-\alpha)^k}, \tag{1}$$

$$U_{\mu}^{(k)}(m, \alpha) = \sum_{n=-\infty}^{\infty} \frac{J_n(m)\, J_{\mu-n}(m)}{(n-\alpha)^k}. \tag{2}$$

Here k is a positive integer, μ an integer, m a real number, α an arbitrary (real or complex) number, and $J_n(m)$ [as well as $J_{n-\mu}(m)$ and $J_{\mu-n}(m)$] is a Bessel (cylindrical) function of the first kind. The use of functions (1) and (2) for the design of an FM detector is discussed in [1].

We have derived several formulas (see [1]) to facilitate the tabulation of the functions (1) and (2). On the basis of these formulas it is sufficient to consider only values of the index μ, equal to zero and one. The quantity m can be assumed to be positive. The real part of number α can be assumed to lie between zero and two. The imaginary part of α can be assumed negative. Thus, the totality of the values of α, for which the functions (1) and (2) should be tabulated, is defined geometrically by the region of the complex-variable plane shown in the figure.



In view of the fact that this region extends to infinity, we must have asymptotic formulas for a more or less simple calculation of the values of the functions (1) and (2) when the imaginary part of α is negative and quite large in absolute magnitude. Formulas of this kind are given in [1].

In the present article we derive asymptotic formulas of different form, more convenient for the calculations.

We assume that $|\alpha|$ is a very large number; this assumption is indeed what takes place in cases of practical interest. In addition, we assume that $\mathrm{Im}(\alpha) < 0$. The latter assumption is indeed no limitation, inasmuch as (see [1])

$$T_{\mu}^{(k)}(m, -\alpha) = (-1)^{k+\mu}\, T_{\mu}^{(k)}(m, \alpha + \mu),$$
$$U_{\mu}^{(k)}(m, -\alpha) = (-1)^{k}\, U_{\mu}^{(k)}(m, \alpha + m).$$

We thus assume that $\alpha = \sigma - i\psi$, where $\psi > 0$. We start from an integral representation of the function (1)

$$T_{\mu}^{(k)}(m, \alpha) = (-1)^{\mu}\, 2^k i^{\mu+k} \sum_{\beta=1}^{k} (-1)^{\beta-1} \frac{\pi^{k-\beta}}{\beta!\,(k-\beta)!}\, I_{\beta-1} \sum_{\lambda=1}^{k+1-\beta} \frac{a_k - \beta^{\lambda}}{(e^{-2\pi i\alpha} - 1)^{\lambda}},$$

where

$$a_{s,\lambda} = \sum_{\gamma=0}^{\lambda-1} (-1)^{\gamma} C_{\lambda-1}^{\gamma} (\lambda - \gamma)^{s}$$

($C_{\lambda-1}^{\gamma}$ denotes the binomial coefficient) and

$$I_s = \int_0^{\pi} Z^s e^{i(\mu-2\alpha)z} J_{\mu}(2m \sin z)\, dZ.$$

We assume that $|\alpha|$ is a very large number. Obviously

$$I_s = \int_0^{\pi} u e^{-2i\alpha z}\, dz,$$

where

$$u = z^s e^{i\mu z} J_{\mu}(2m \sin z).$$

Integrating by parts we get

$$I_s = -\sum_{\delta=0}^{n} \frac{1}{(2i\alpha)^{\delta+1}} \left. \frac{d^{\delta} u}{dz^{\delta}} e^{-2i\alpha z} \right|_0^{\pi} + \frac{1}{(2i\alpha)^{n+1}} \int_0^{\pi} \frac{d^{n+1} u}{dz^{n+1}} e^{-2i\alpha z}\, dz.$$

As already mentioned, we can assume that $\mu \geq 0$. In this case, when $\nu < \mu$, we have

$$\left. \frac{d^{\nu} u}{dz^{\nu}} e^{-2i\alpha z} \right|_0^{\pi} = 0.$$

This result is obtained in the following fashion. By the Leibnitz rule

$$\frac{d^{\nu} u}{dz^{\nu}} = \sum_{r=0}^{\nu} A_r(z) \frac{d^r J_{\mu}(2m \sin z)}{dz^r},$$

and none of the functions $A_r(z)$ has singularities at finite values of z. Further

$$\frac{d^r J_{\mu}(2m \sin z)}{dz^r} = \sum_{k=1}^{r} B_k(z) \frac{d^k J_{\mu}(\lambda)}{d\lambda^k},$$

where $\gamma = 2m \sin z$; the functions $B_k(z)$ also have no singularities at finite values of z. Consequently

$$\frac{d^{\nu} u}{dz^{\nu}} = \sum_{r=0}^{\nu} C_r(z) \frac{d^r J_{\mu}(\lambda)}{d\lambda^r},$$

where the functions $C_r(z)$ have no singularities for any finite values of z.

From the formula

$$\frac{d^{\nu} J_{\mu}(\lambda)}{d\lambda^{\nu}} = \frac{1}{2^{\nu}} \sum_{n=0}^{\nu} (-1)^h C_{\nu}^h J_{\mu-\nu+2h}(\lambda)$$

it follows that

$$\frac{d^\nu u}{dz^\nu} = \sum_{r=0}^{\nu} D_r(z)\, J_{\mu-\nu+2r}(2m \sin z),$$

with the functions $D_r(z)$ having no singularities for any finite values of z. But since

$$J_n(0) = \begin{cases} 1 \text{ when } n = 0, \\ 0 \text{ when } n > 0, \end{cases}$$

we have

$$\left. \frac{d^\nu u}{dz^\nu} e^{-2i\alpha z} \right|_0^\pi = 0 \; (\nu < \mu), \qquad \left. \frac{d^\mu u}{dz^\mu} e^{-2i\alpha z} = \right|_0^\pi D_0(z)\, e^{-i\alpha z} \Big|_0^\pi .$$

Let us calculate $D_0(z)$. Obviously

$$\frac{d^\mu u}{dz^\mu} = \sum_{r=0}^{\mu-1} A_r(z) \frac{d^2 J_\mu(2m \sin z)}{dz^2} + z^s e^{i\mu z} \frac{d^\mu z_\mu(2m \sin z)}{dz^\mu}.$$

Furthermore

$$\frac{d^\mu J_\mu(2m \sin z)}{dz^\mu} = \sum_{k=1}^{\mu-1} B_k(z) \frac{d^k J_\mu(\lambda)}{d\lambda^\mu} + (2m \cos z)^\mu \frac{d^\mu J_\mu(\lambda)}{d\lambda^\mu}.$$

Consequently,

$$\frac{d^\mu u}{dz^\mu} = \sum_{r=0}^{\mu-1} C_r(z) \frac{d^r J_\mu(\lambda)}{d\lambda^r} + z^s e^{i\mu z} (2m \cos z)^\mu \frac{d^\mu J_\mu(\lambda)}{dz^\mu}.$$

We find that

$$D_0(z) = \frac{1}{2^\mu} z^s e^{i\mu z} (2m \cos z)^\mu.$$

Thus

$$D_0(z)\, e^{-2i\alpha z} \Big|_0^\pi = m^\mu (e^{-2\pi i\alpha} - 1),$$

if s = 0 and

$$D_0(z)\, e^{-2i\alpha z} \Big|_0^\pi = \pi^s m^\mu e^{-2\pi i\alpha},$$

if s > 0. From this it follows that

$$I_0 = -\frac{1}{(2i\alpha)^{\mu+1}} m^\mu (e^{-2\pi i\alpha} - 1) - \sum_{\delta=\mu+1}^{n} \frac{1}{(2i\alpha)^{\delta+1}} \left. \frac{d^\delta u}{dz^\delta} e^{-2i\alpha z} \right|_0^\pi + \\ + \frac{1}{(2i\alpha)^{n+1}} \int_0^\pi \frac{d^{n+1} u}{dz^{n+1}} e^{-2i\alpha z}\, dz.$$

For the integral in the right half of this formula we can write

$$\left|\int_0^\pi \frac{d^{n+1}u}{dz^{n+1}}\, e^{-2i\alpha z}\, dz\right| = \left|\int_0^\pi \frac{d^{n+1}u}{dz^{n+1}}\, e^{-2i\alpha z}\, e^{-2\psi z}\, dz\right| < \int_0^\pi \left|\frac{d^{n+1}}{dz^{n+1}}\right| dz.$$

The integral of the right half of this formula is independent of α. Consequently, for all values of μ, m, and s the expression

$$\frac{1}{(2i\alpha)^{n+1}} \int_0^\pi \frac{d^{n+1}u}{dz^{n+1}}\, e^{-2i\alpha z}\, dz$$

can be made as small as desired in absolute magnitude, provided we take $|\alpha|$ sufficiently large. We thus arrive at the asymptotic formula

$$I_0 \sim -\frac{1}{(2i\alpha)^{\mu+1}}\, m^{\mu} \left(e^{-2\pi i\alpha} - 1\right) - \frac{1}{(2i\alpha)^{\mu+2}} \left. \frac{d^{\mu+1}u}{dz^{\mu+1}}\, e^{-2i\alpha z} \right|_0^\pi -$$
$$- \frac{1}{(2i\alpha)^{\mu+3}} \left. \frac{d^{\mu+2}u}{dz^{\mu+2}}\, e^{-2i\alpha z} \right|_0^\pi - \ldots - \frac{1}{(2i\alpha)^{n+1}} \left. \frac{d^{n}u}{dz^{n}}\, e^{-2i\alpha z} \right|_0^\pi - \ldots .$$

In the case when s > 0 we obtain quite analogously the formula

$$I_s \sim -\frac{1}{(2i\alpha)^{\mu+1}}\, \pi^s m^{\mu} e^{-2\pi i\alpha} - \frac{1}{(2i\alpha)^{\mu+2}} \left. \frac{d^{\mu+1}}{dz^{\mu+1}}\, e^{-2i\alpha z} \right|_0^\pi -$$
$$- \frac{1}{(2i\alpha)^{\mu+3}} \left. \frac{d^{\mu+2}u}{dz^{\mu+3}}\, e^{-2i\alpha z} \right|_0^\pi - \ldots - \frac{1}{(2i\alpha)^{n+1}} \left. \frac{d^{n}u}{dz^{n}}\, e^{-2i\alpha z} \right|_0^\pi .$$

Let us proceed to derive analogous asymptotic formulas for the function (2). We start in this derivation from the integral representation of the function

$$U_{\mu}^{(k)}(m,\alpha) = (-1)^{\mu}\, 2^k\, i^k \sum_{\beta=1}^{k} (1)^{\beta-1} \frac{\pi^{k-\beta}}{\beta!\,(k-\beta)!}\, K_{\beta-1} \sum_{\lambda=1}^{k-\beta+1} \frac{a_k - \beta_1\lambda}{(e^{-2\pi i\alpha}-1)\lambda},$$

where

$$K_s = \int_0^\pi z^s e^{i(\mu-2\alpha)z} J_{\mu}(2m\cos z)\, dz.$$

It is obvious that

$$K_s = \int_0^\pi v e^{-2i\alpha z}\, dz,$$

where

$$v = z^s e^{i\mu z} J_{\mu}(2m\cos z).$$

Integrating by parts we obtain

$$K_s = -\sum_{\delta=0}^{n} \frac{1}{(2i\alpha)^{\delta+1}} \left. \frac{d^{\delta}v}{dz^{\delta}}\, e^{-2i\alpha z} \right|_0^\pi + \frac{1}{(2i\alpha)^{n+1}} \int_0^\pi \frac{d^{n+1}v}{dz^{n+1}}\, e^{-2i\alpha z}\, dz.$$

Let us calculate the first term in the right half of this formula. Obviously

$$\Big|_0^{\pi} v e^{-2i\alpha z} = \Big|_0^{\pi} z^s e^{i(\mu-2\alpha)z} J_\mu (2m \cos z) = J_\mu (2m)\,(e^{-2\pi i\alpha} - 1),$$

if s = 0 and

$$\Big|_0^{\pi} v e^{-2i\alpha z} = \pi^s J_\mu (2m)\, e^{-2\pi i\alpha},$$

if s > 0. The term containing the integral in the right half of the formula for K_s can be made as small as desired in absolute magnitude, provided we take $|\alpha|$ sufficiently large. We obtain, consequently, the following asymptotic formulas (s > 0)

$$K_0 \sim -\frac{J_\mu (2m)}{2i\alpha}(e^{-2\pi i\alpha} - 1) - \frac{1}{(2i\alpha)^2}\Big|_0^{\pi} \frac{dv}{dz} e^{-2i\alpha z} -$$

$$-\frac{1}{(2i\alpha)^3}\Big|_0^{\pi} \frac{d^2 v}{dz^2} e^{-2i\alpha z} - \ldots - \frac{1}{(2i\alpha)^{n+1}}\Big|_0^{\pi} \frac{d^n v}{dz^n} e^{-2i\alpha z} - \ldots$$

$$K_s \sim -\frac{\pi^s J_\mu (2m)}{2i\alpha} e^{-2\pi i\alpha} - \frac{1}{(2i\alpha)^2}\Big|_0^{\pi} \frac{dv}{dz} e^{-2i\alpha z} -$$

$$-\frac{1}{(2i\alpha)^3}\Big|_0^{\pi} \frac{d^2 v}{dz^2} e^{-2i\alpha z} - \ldots - \frac{1}{(2i\alpha)^{n+1}}\Big|_0^{\pi} \frac{d^n v}{dz^n} e^{-2i\alpha z} - \ldots$$

If we confine ourselves in the asymptotic expansions to the initial terms, we obtain the approximate formulas (s > 0)

$$I_0 \simeq \frac{1}{(2i\alpha)^{\mu+1}} m^\mu (e^{-2\pi i\alpha} - 1), \quad I_s \simeq \frac{1}{(2i\alpha)^{\mu+1}} \pi^s m^\mu e^{-2\pi i\alpha},$$

$$K_0 \simeq \frac{J_\mu (2m)}{2i\alpha}(e^{-2\pi i\alpha} - 1), \; K_s \simeq \frac{\pi^s J_\mu (2m)}{2i\alpha} e^{-2\pi i\alpha}.$$

The formulas obtained simplify the calculation of the result of the detection of frequency modulated oscillations, compared with the formulas given in [1].

REFERENCES

1. V.K. Turkin, G.A. Levin, Doklady AN SSSR, 1956, 106, 6.

Received by editors 9 May 1961

ON THE THEORY OF DETECTION OF COMPLEX FREQUENCY MODULATED OSCILLATIONS

G.A. Levin, V.K. Turkin

A theory is given for the detection of frequency modulated oscillations for the case when the modulating function is an arbitrary periodic function of the time. The results of reference [1] are generalized.

In the present article we consider the detection of complicated frequency-modulated oscillations of the form

$$u = A \cos\,[\omega_0 t + f\,(t)], \tag{1}$$

where f(t) is a periodic function. Replacing f(t) by a segment of its Fourier series, we can write for Eq. (1) approximately

$$u = A \cos\Big[\omega_0 t + \sum_{k=1}^{s} (m_k \sin k\Omega t + m_k' \cos k\Omega t)\,\Big]. \tag{2}$$

We shall henceforth find it more convenient to use the complex notation; we therefore use in lieu of (2)

$$u = A \exp i\Big[\omega_0 t + \sum_{k=1}^{s} (m_k \sin k\Omega t + m_k' \cos k\Omega t)\Big]. \tag{2'}$$

Expression (2') can be broken up into harmonics, the amplitudes of which are expressed in terms of hypercylindrical functions (Bessel functions of several variables)

$$u = A \sum_{n=-\infty}^{\infty} J_n (m_1, m_2, \ldots, m_s;\ m_1', m_2', \ldots, m_s')\, e^{i(\omega_0 + n\Omega)t}, \tag{3}$$

where

$$J_n\,(m_1, m_2, \ldots, m_s;\ m_1', m_2', \ldots, m_s') =$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp i\Big[\sum_{r=1}^{s} (m_r \sin r\varphi + m_r' \cos r\varphi) - n r\Big] d\varphi.$$

We shall henceforth denote this hypercylindrical function briefly by $J_n(m_r, m'_r)$.

We assume that the complex frequency-modulated oscillation (2') is applied to a frequency detector, which at first supplements the frequency modulation with amplitude modulation, and then subjects the result to amplitude detection. Amplitude modulation converts (3) into

$$v = A \sum_{n=-\infty}^{\infty} b_n J_n (m_r, m_r')\, e^{i(\omega_0 + n\Omega)t}, \tag{4}$$

where b_n is the coefficient of amplitude modulation (in general, complex).

The frequency and amplitude modulated oscillation (4) is applied to an amplitude detector; if the characteristic of the latter has the form $I = f(v)$, then

$$I = f\Big[A \sum_{n=-\infty}^{\infty} b_n J_n(m_r, m_r') \, e^{i(\omega_0+n\Omega)t}. \tag{5}$$

In the case of a quadratic characteristic $f(v) = \gamma_0 + \gamma v^2$, we obtain the formula

$$I = \gamma_0 + \frac{\gamma A^2}{2} \sum_{\mu=-\infty}^{\infty} \Big[\sum_{n=-\infty}^{\infty} b_n b_{n-\mu} J_n(m_r, m_r') J_{n-\mu}(m_r, m_r')\Big] e^{i\mu\Omega t} + \\ + \frac{\gamma A^2}{2} \sum_{\mu=-\infty}^{\infty} \Big[\sum_{n=-\infty}^{\infty} b_n b_{\mu-n} J_n(m_r, m_r') J_{\mu-n}(m_r, m_r')\Big] e^{i(2\omega_0+\mu\Omega)t}. \tag{6}$$

Consequently, when expanding (6) into harmonics it is necessary to calculate the expressions

$$\sum_{n=-\infty}^{\infty} b_n b_{n-\mu} J_n(m_r, m_r') J_{n-\mu}(m_r, m_r'), \tag{7}$$

$$\sum_{n=-\infty}^{\infty} b_n b_{\mu-n} J_n(m_r, m_r') J_{\mu-n}(m_r, m_r'). \tag{8}$$

In the calculation of expressions (7) and (8) it is necessary to know the dependence of the coefficient of amplitude modulation b_n on the number of the harmonic n. The greatest theoretical and practical interest is attached to the case when b_n is a rational-fraction function of n (in most cases this rational fraction can be considered to be regular). This case is encountered whenever one deals with amplitude modulation with the aid of an electric network, described by a system of linear differential equations with constant coefficients (we are considering steady-state oscillations). In this case the product $b_n b_{n-\mu}$ and $b_n b_{\mu-n}$ can be broken up into terms of the form $c/(n - \alpha)^k$, where k is a positive integer; α is the root of the polynomial which is the denominator of the regular rational fraction under consideration, and c is an arbitrary number (k, α, and c are independent of n). Consequently, the expression (7) can be broken up into terms of the form

$$C \sum_{n=-\infty}^{\infty} \frac{1}{(n-\alpha)^k} J_n(m_r, m_r') J_{n-\mu}(m_r, m_r'),$$

where C is independent of n. Analogously, expression (8) can be broken up into terms of the form

$$D \sum_{n=-\infty}^{\infty} \frac{1}{(n-\alpha)^k} J_n(m_r, m_r') J_{\mu-n}(m_r, m_r'),$$

where D is independent of n.

Thus, we must find a method for calculating the expressions

$$T_\mu^{(k)}(m_r, m_r', \alpha) = \sum_{n=-\infty}^{\infty} \frac{1}{(n-\alpha)^k} J_n(m_r, m_r') J_{n-\mu}(m_r, m_r'), \tag{9}$$

$$U_\mu^{(k)}(m_r, m_r', \alpha) = \sum_{n=-\infty}^{\infty} \frac{1}{(n-\alpha)^k} J_n(m_r, m_r') J_{\mu-n}(m_r, m_r'). \tag{10}$$

To calculate expressions (9) and (10) one can use the theorem for addition of hypercylindrical functions, which can be represented in the form:

$$\sum_{n=-\infty}^{\infty} e^{in\rho} J_n(m_r, m'_r) J_{n-\mu}(p_r, p'_r) = \\ = J_\mu(m_r \cos r\rho - m'_r \sin r\rho - p_r, m_r \sin r\rho + m'_r \cos r\rho + p'_r). \tag{11}$$

By integrating with respect to ρ we obtain from (11) an integral representation of the function $T_\mu^{(1)}(m_r, m'_r, \alpha)$:

$$T_\mu^{(1)}(m_r, m'_r, \alpha) = \frac{1}{e^{-2\pi i\alpha} - 1} \int_0^{2\pi} e^{-i\alpha\rho} J_\mu(m_r \cos r\rho - m'_r \sin r\rho - \\ - m_r, m_r \sin r\rho + m'_r \cos r\rho + m'_r)\, d\rho. \tag{12}$$

The integral representation for $T_\mu^{(k)}(m_r, m'_r, \alpha)$ when $K > 1$ is obtained from (11) by differentiating with respect to α.

Let $|\alpha|$ be quite large (this is precisely the case of practical interest). We then obtain from the integral representations for the functions $T_\mu^{(k)}(m_r, m'_r, \alpha)$, by integrating by parts, asymptotic expansions for these functions. If we take only the initial term in the asymptotic expansion obtained from (11), we obtain an approximate formula (which is suitable when $|\alpha| \gg 1$)

$$T_\mu^{(1)}(m_r, m'_r, \alpha) \simeq \frac{i^\mu}{\alpha} J_\mu(0, 2m'_r). \tag{13}$$

Formula (13) holds true if $\mathrm{Im}(\alpha) < 0$; this condition is essentially not a limitation, for the case $\mathrm{Im}(\alpha) > 0$ can be reduced to the case $\mathrm{Im}(\alpha) < 0$ with the aid of formula

$$T_\mu^{(k)}(m_r, m'_r, \alpha) = (-1)^k T_\mu^{(k)}(-m_r, m'_r, \alpha).$$

The values of the hypercylindrical function can be calculated relatively easily with the aid of a series obtained from the foregoing integral representation of this function.

The integral representations and the asymptotic expansions for the functions (10) can be derived analogously. It is necessary to start here from the theorem for the addition of hypercylindrical functions, written in the form

$$\sum_{n=-\infty}^{\infty} e^{in\rho} J_n(m_r, m'_r) J_{\mu-n}(p_r, p'_r) = \\ = J_\mu(m_r \cos r\rho - m'_r \sin r\rho + p_r + m_r \sin r\rho + m'_r \cos r\rho + p'_r).$$

The formulas obtained make it possible to calculate in practice the detection of frequency-modulated oscillations in the case when the modulating signal is an arbitrary periodic function of the time.

REFERENCES

1. V.K. Turkin, G.A. Levin, Doklady AN SSSR, 1956, 106, 6.

Received by editor 9 May 1961

AUTOMATIC PHASE CONTROL WITH DELAY

V.A. Shpolyanskiy, L.S. Tyufyakin,* P.P. Korsakov

An automatic phase control system is considered with a bandpass amplifier in the control loop. The special operating features of such a control are due to the presence of additional delay in the loop. The stability, the interference immunity, and the quality of the regulation in such a phase control system are investigated, and the possibility of improving its dynamic and filtering properties by introducing a derivative control is analyzed.

INTRODUCTION

A system of automatic phase control of the frequency, with a bandpass amplifier tuned to some intermediate frequency included in the control loop (Fig. 1), has been in extensive use in recent times. The need for such an amplifier arises as a rule at centimeter wavelengths, when the reference signals are weak oscillations of molecular generators or higher harmonics (usually also weak) of crystal-stabilized generators. Analogous systems are used also in radio receivers when the automatic phase control of the system is based on the received signal [1-5].

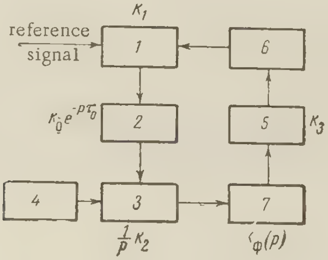


Fig. 1

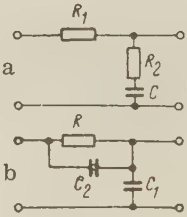


Fig. 2

Fig. 1. Block diagram of automatic phase control system:
1—mixer; 2—intermediate frequency amplifier; 3—phase detector; 4—standard generator for intermediate frequency oscillation; 5—controlling element; 6—tuned generator; 7—low-pass filter.

Fig. 2. Proportional-integrating filter circuits $\left(K_f(p) = \frac{T_2 p + 1}{T_1 p + 1}\right)$: a— $T_1 = (R_1 + R_2)C$, $T_2 = R_2 C$;

b— $T_1 = (C_1 + C_2)R$, $T_2 = RC_2$.

The presence of an amplifier in the automatic-frequency-control loop changes in many cases (primarily as a result of the delay produced) the properties of the system and makes it necessary to take these changes into account. The effect of delay on a phase control system has been the subject of several papers [1, 3, 4].* However, discussion of the dynamics of systems with delay in the published literature have been limited to an investigation of the stability and of the locking band of the simplest system of automatic phase control with integrating RC filter, which, as will be shown below, has the worst characteristics if the delay is to any degree appreciable.

---

*Deceased.

One way of improving the dynamic characteristics of delay systems is, as is well known, the introduction of derivative control. The use of the combined frequency-phase automatic control for this purpose results in appreciable benefits [5]. An analogous effect can be attained also by simpler means, introducing a derivative control into the automatic phase control system, with the aid of a proportional-integrating filter [3] located past the phase detector (Fig. 2). The present article is devoted to an analysis of the dynamic and filtering properties of such a system.

1. TRANSFER FUNCTION AND STABILITY OF THE AUTOMATIC PHASE CONTROL SYSTEM

If the transfer function of a multi-stage passband amplifier in its usual form [11] is taken into account in the initial differential equation for automatic phase control, the order of this differential equation is appreciably increased, so that the subsequent investigation becomes quite difficult. To simplify the problem somewhat we can take into account the time delay introduced into the automatic phase control system by the IF stages, by replacing the IF amplifier with a pure-delay link with a transfer function in operator form $K_{\tau_3} = K_0 e^{-\tau_3 p}$, where τ_3 is connected with the passband of the n-stage amplifier $\Delta f_{0.7}$ by the well known expression

$$\tau_3 = \frac{n}{\pi \Delta f_{0.7}} \sqrt{\sqrt[n]{2} - 1}\,; \tag{1}$$

K_0 is the maximum gain. The accuracy of this approximation, in terms of the frequency characteristic, does not exceed 10% even when n > 3.

To determine the stability and the filtering characteristics of the automatic phase control system, we make use of the operator transfer function W(p) of the system, linearized with respect to small deviations (from the synchronism position) due to random changes in the frequency of the tuned generator or due to fluctuation noise.

Recognizing that

$$W(p) = \frac{\Delta\omega_n(p)}{\Delta\omega_0(p)}, \tag{2}$$

where $\Delta\omega_0(p)$ is the variation of the frequency of the tuned generator (initial signal); $\Delta\omega_n(p)$ is the variation in the frequency of the tuned generator due to the automatic phase control system (response), we obtain, by virtue of the structural diagram (Fig. 1)

$$W(p) = \frac{1}{1 + \dfrac{p e^{-\tau_3 p}}{\Delta\omega_h K_f(p)}}\,; \tag{3}$$

$\Delta\omega_h$ is the holding band of the system, equal to the product of the maximum transfer function of all the system elements in the synchronism mode; $K_f(p)$ is the normalized transfer function of the low-pass filter.

Substituting in (3) the operator transfer function of the proportional-integrating filter and going to dimensionless parameters, we get

$$W(p_1) = \frac{\Delta_2 p_1 + 1}{\Delta_1 p_1^2 e^{\Delta_3 p_1} + p_1 (\Delta_2 + e^{\Delta_3 p_1}) + 1}, \tag{4}$$

---

*After submitting this paper to the editor, an interesting paper appeared [10], devoted to an investigation of the stability of an automatic phase control system for an arbitrary low-pass filter circuit.

where

$$\begin{aligned} p_1 &= \frac{p}{\Delta\omega_y} = p\tau; \\ \tau &= \frac{1}{\Delta\omega_y}; \\ \Delta_1 &= \frac{T_1}{\tau}; \\ \Delta_2 &= \frac{T_2}{\tau}; \\ \Delta_3 &= \frac{T_3}{\tau}. \end{aligned} \tag{4a}$$

Using the Mikhailov criterion, we find that the automatic control mode is stable if the delay in the system does not exceed

$$\Delta_{3\,\mathrm{cr}} = \frac{\frac{\pi}{2} - \operatorname{arc\,tg} \beta(\Delta_1; \Delta_2) + \operatorname{arc\,tg} \eta\beta(\Delta_1; \Delta_2)}{\beta(\Delta_1; \Delta_2)}, \tag{5}$$

where

$$\beta(\Delta_1; \Delta_2) = \sqrt{\frac{\sqrt{(1-\Delta_2^2)^2 + 4\Delta_1^2} - (1-\Delta_2^2)}{2}};$$

$$\eta = \frac{\Delta_2}{\Delta_1}.$$

In the case of an integrating filter ($\Delta_2 = 0$), expression (5) yields the well known value for the critical delay time in an automatic phase control system, and the stability conditions [1, 3, 5]:

$$\Delta_3 \leqslant \Delta_{3\,\mathrm{cr}} = \frac{\frac{\pi}{2} - \operatorname{arc\,tg} \sqrt{\frac{\sqrt{1+4\Delta_1^2} - 1}{2}}}{\sqrt{\frac{\sqrt{1-4\Delta_1^2} - 1}{2}}} \Delta_1, \tag{6}$$

$$\text{or} \quad \Delta_3 \leqslant \Delta'_{3\,\mathrm{cr}} = 1 \tag{6a}$$

When $\Delta_1 \gg 1$. An analogous limiting expression can be readily obtained also for a system with proportional-integrating filter with sufficiently large time constants $\Delta_1 \gg 1$ and $\Delta_2 \gg 1$. Here Eq. (5) simplifies to the form

$$\Delta_{3\,\mathrm{cr}} = \Delta'_{3\,\mathrm{cr}} + \frac{\pi}{2}\frac{1}{\eta}, \tag{7}$$

or, by virtue of (6a),

$$\Delta_{3\,\mathrm{cr}} = 1 + \frac{\pi}{2}\frac{1}{\eta}. \tag{7a}$$

It is easy to see that the use of a proportional-integrating filter increases the value of the critical delay time compared with its value in a system with integrating filter.

The gain in $\Delta_{3\mathrm{cr}}$ is determined here, for sufficiently large $\Delta_1 \gg 1$ and $\Delta_2 \gg 1$, by the magnitude of their ratio η, and if sufficiently small values of η are chosen this gain may prove to be considerable.

To estimate the influence of the proportional-integrating filter on the critical delay, we can use the plot of Fig. 3, based on formula (5), for several fixed values of η.

In practice, interest frequently attaches to the dependence of the critical value of the passband of the n-stage IF amplifier (i.e., the band at which Δ_3, introduced by the IF

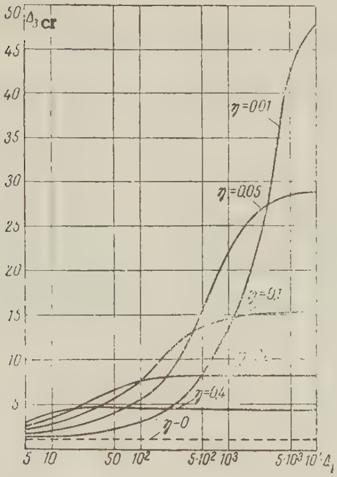


Fig. 3

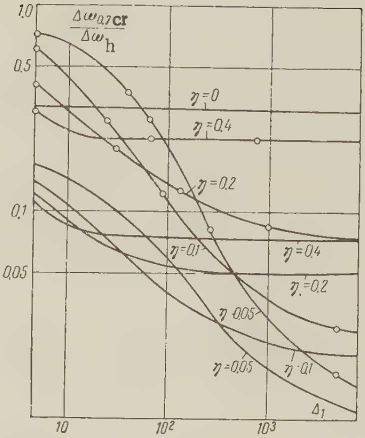


Fig. 4

Fig. 3. Dependence of critical delay time on the filter parameters Δ_1 and η.

Fig. 4. Dependence of the critical band of the IF amplifier on the filter parameters Δ_1 and η for an n-stage tuned IF amplifier. Solid curves — n = 1; circles — n = 10.

amplifier, is equal to $\Delta_{3\mathrm{cr}}$) on the parameters of the low-pass filter.

This dependence, plotted in Fig. 4 with account taken of (1) and (5), shows that the minimum permissible passband, from the point of view of stability, can be considerably less than the holding band of the system.

This circumstance, as will be shown below, makes it possible to increase the interference immunity in many applications of automatic phase control, by suitable choice of the system parameters.

2. INTERFERENCE IMMUNITY OF AUTOMATIC PHASE CONTROL SYSTEM

As is well known, one of the factors that limits the accuracy of an automatic phase control system is the effect of fluctuation noise, which causes random variations of the phase of the tuned generator.

The effect of this action has been calculated, for automatic phase control systems without delay, in many papers [3, 6, 7] under the assumption that the noise is white. However, the presence of delay prevents us in some cases from using the deductions of these papers concerning the filtering properties of the automatic phase control.

In addition, in many automatic control applications where initial synchronization is possible, the IF bandwidth can be reduced to values considerably less than the holding bandwidth of the system, so that it becomes necessary to take into account the influence of the frequency characteristic of the IF amplifier on the input-noise spectrum.

Let us consider the filtering properties of automatic phase control under the aforementioned circumstances, assuming that the noise is a fluctuating voltage generated in the mixer—IF amplifier system. We are interested here only in tuned-oscillator phase fluctuations due to changes which this noise introduces into the phase of the incoming signal.

Assuming the noise level to be sufficiently small compared with the threshold signal, we can assume that the dispersion of the phase of the signal at the input of the phase detector is given by the expression

$$\overline{\delta\theta^2} = \frac{\overline{U}_{\mathrm{n}}^2}{U_{\mathrm{sIF}}^2} = \frac{P_{\mathrm{n}}}{P_{\mathrm{sIF}}}, \tag{8}$$

where $\overline{U}_n$, P_n, U_{sIF}, and P_{sIF} are respectively the mean-square values of the noise and signal at the output of the IF amplifier.

To determine the dispersion of the noise voltage of the mixer — IF amplifier system, we shall assume this noise to be uncorrelated, and the IF amplifier to be an ordinary real amplifier (and not a pure delay element). This is necessary because a pure-delay element has an infinitely broad passband and such an idealization would lead to an infinitely large value of $\overline{U^2_n}$. On the other hand, this assumption is possible by virtue of the fact that the automatic phase control system is not closed by a feedback loop with respect to the fluctuations of the signal voltage, and the transfer function from the input of the IF amplifier to the phase detector is determined only by the transfer function N(p) of the IF amplifier. At the same time, with respect to the fluctuations of the phase, the automatic phase control system is closed and is described by the transfer function (3), in which the IF amplifier is regarded as a pure-delay element.

Taking the foregoing into consideration and using the well known relations for the mixer—IF amplifier system with the aforementioned transfer function N(p), we can rewrite Eq. (8) for an input signal of power P_s in the form

$$\overline{\delta\theta^2} = \frac{kTF\int\limits_0^\infty |N(j\omega)|^2\,d\omega}{2\pi P_s} \tag{9}$$

(F is the noise factor, k is Boltzmann's constant, and T is the absolute temperature), and we obtain for the spectral density at the input of the IF amplifier the expression

$$\Omega_{in}(\omega) = \frac{kTF\,|N(j\omega)|^2}{2\pi P_s}\,. \tag{10}$$

Multiplying the last expression by the square of the modulus of the frequency characteristic of the closed automatic phase control system (expression (3) with $p = j\omega$) we obtain the following expression for the spectral density of the phase fluctuations of the tuned generator:

$$\Omega_{out}(\omega) = \frac{kTF\,|N(j\omega)|^2}{2\pi P_c}\,|W(j\omega)|^2 \tag{11}$$

and for the phase dispersion

$$\Delta\theta^2 = \frac{kTF}{2\pi P_s}\Pi_n, \tag{12}$$

where

$$\Pi_n = \int\limits_0^\infty |N(j\omega)|^2\,|W(j\omega)|^2\,d\omega. \tag{13}$$

Thus, at specified values of the signal power and at specified mixer and IF amplifier parameters, the filtering properties of the automatic phase control system are determined completely by the noise band, i.e., by expression (13).

Let us consider first the effect of the parameters of the low-pass filter and the delay time τ_3 on the value of π_n in the case when the IF bandwidth is much broader than the passband of the closed-loop automatic phase control system ($|N(j\omega)|^2 \equiv 1$).

If the system includes an integrating low-pass filter, we can obtain [8], by means of (3), an expression for π_n:

$$\Pi_{in} = \frac{1}{\tau_0}\int\limits_0^\infty \frac{d\xi}{1 + 2\xi\sin\Delta_3\xi + (1 - 2\Delta_1\xi\cos\Delta_3\xi)\,\xi^2 + \Delta_1^2\,\xi^4}\,, \tag{14}$$

where ξ is the dimensionless quantity, $\xi = \omega/\Delta\omega_h = \omega\tau$.

In the absence of delay ($\Delta_3 = 0$), expression (14) yields the well known [3] relation for the automatic phase system with integrating filter

$$\Pi_n = \frac{1}{\tau}\int_0^\infty \frac{d\xi}{1-(1-2\Delta_1)\xi^2+\Delta_1^2\xi^4} = \frac{\pi}{\tau}. \tag{15}$$

To estimate the effect of delay on the noise band of the system, it is necessary to evaluate the integral (14). Since this is in general difficult to do, the value of the integral for different Δ_1 were obtained numerically [8] by graphic integration.

The dependence of the relative value of the noise band ξ on the delay Δ_3 and the parameters of the low-pass filter

$$\xi = \frac{\Pi_n}{\Pi_{n0}} = \frac{\tau}{\pi}\Pi_n,$$

calculated by formula (14) and shown in Fig. 5, lead to the following conclusions.

In the presence of delay in the automatic phase control channel with integrating filter, the noise band depends appreciably on the constant of the low-pass filter Δ_1, and increases with it; the dependence of π_n on the constant Δ_1 becomes sharply stronger with increasing delay time.

Thus, in an automatic phase control system with integrating low-pass filter, we have in many cases $\pi_n \gg \pi/\tau$, and the greater the delay time and the greater the constant Δ_1, the stronger this inequality becomes.

Taking the foregoing into account, it is necessary to exercise particular care, when estimating the interference immunity of the system, in the application of the results of many investigations on the filtering properties of the automatic phase control, for in many cases these results are valid only for systems without delay. An analogous effect is observed also in the case of automatic phase control with a proportional integrating filter. The noise band of such a system is determined by the expression

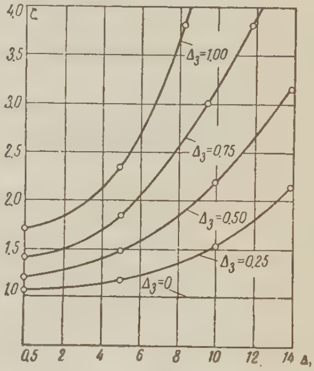


Fig.5. Dependence of the relative noise bandwidth of the system with integrating low-pass filter on the filter constant Δ_1. Solid line — system without delay; circles — system with delay.

$$\Pi_n^* = \frac{1}{\tau}\int_0^\infty \frac{(1+\Delta_2^2\xi^2)\,d\xi}{1+(1+\Delta_2^2)\xi^2+\Delta_1^2\xi^4+2\xi^2(\Delta_2-\Delta_1)\cos\Delta_3\xi-2\xi(1+\Delta_1\Delta_2\xi)\sin\Delta_3\xi}, \tag{16}$$

which is derived in analogy with the foregoing. In the absence of delay we have the well known expression for the noise band [3] in a system with proportional-integrating filter:

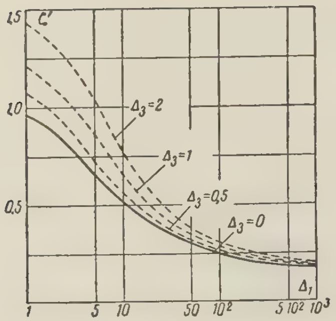
$$\Pi'_{n0} = \frac{1}{\tau}\int_0^\infty \frac{(1+\Delta_2^2\xi^2)\,d\xi}{1+[(1+\Delta_2)^2-2\Delta_1]\xi^2+\Delta_1^2\xi^4}. \tag{17}$$

The results of numerical integration of expression (16) for two values of η (0.4 and 0.2) for different values of the delay are shown in Figs. 6 and 7. As in the case of an integrating filter, the relative noise band is

$$\xi^* = \frac{\Pi_n^*}{\Pi_{n0}} = \frac{\tau}{\pi}\Pi_n$$

and increases in the presence of delay in the system, compared with its value in the absence of delay. This effect, however, is much less pronounced in this case, and its influence is all the less, the greater the ratio of the constants η.

At sufficiently large Δ_1 and at a fixed ratio η, the noise band tends to a limiting value



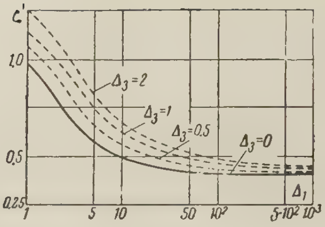


Fig. 6. Dependence of relative noise band of automatic phase control system with proportional-integrating low-pass filter on the parameter Δ_1 for $\eta = 0.2$. Solid curve — system without delay; dashed curves — system with delay.

Fig. 7. Dependence of relative noise bandwidth of automatic phase control with proportional-integrating low-pass filter on the parameter Δ_1 for $\eta = 0.4$. Solid curve — system without delay; dashed curve — system with delay.

$$\Pi^{*}_{n0} = \lim_{\Delta_1, \Delta_2 \to \infty} \Pi^{*}_{n} = \frac{1}{\tau} \int_0^{\infty} \frac{d\xi}{1 + \frac{\xi^2}{\eta} + \frac{2\xi}{\eta} \sin \xi}, \tag{18}$$

which at sufficiently small η assumes the form

$$\Pi^{*}_{n0} = \frac{\pi}{\tau} \eta = \Pi_{n0} \eta. \tag{19}$$

In many cases when automatic phase control is used and where initial synchronization is possible, we encounter the case when the pass-band of the IF amplifier becomes comparable with the passband of the system. Assuming, for the sake of being specific, that the frequency characteristic of the IF amplifier is determined by an n-stage amplifier with circuits tuned to resonance, we obtain

$$[N(j\xi)]^2 = \frac{1}{\left[1 + \frac{2\xi}{\Delta\xi_{0,7}} \frac{1}{\psi(n)}\right]^n}, \tag{20}$$

where

$$\Delta\xi_{0,7} = \frac{\Delta\omega_{0,7}}{\Delta\omega_y}; \quad \psi(n) = \frac{1}{\sqrt{\sqrt[n]{2} - 1}}.$$

Taking (13), (14), and (20) into account, we obtain for the system with integrating low-pass filter the expression

$$\Pi_n = \frac{1}{\tau} \int_0^{\infty} \frac{d\xi}{\left[1 - 2\xi \sin \xi \frac{\alpha_1(n)}{\Delta\xi_{0,7}} + \left(1 - 2\Delta_1 \cos \xi \frac{\alpha_1(n)}{\Delta\xi_{0,7}}\right)^2 \xi + \Delta_1^2 \xi^4\right]\left[1 + \frac{2\xi}{\Delta\xi_{0,7}} \frac{1}{\psi(n)}\right]^n}, \tag{21}$$

where $\alpha_1(n)$ is a coefficient that depends on the type of the amplifier and on the number of stages [11].

By numerically integrating Eq. (21) we can obtain the dependence of the relative noise band of the automatic phase control system on the passband of the IF amplifier for different values of the time constant of the low-pass filter.

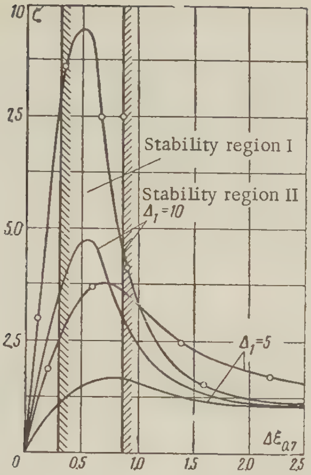


Fig. 8. Dependence of relative noise bandwidth of automatic phase control with integrating low-pass filter on the IF amplifier bandwidth for different values of Δ_1 and n, the number of IF amplifier stages. Solid curves — n = 1; circles — n=10.

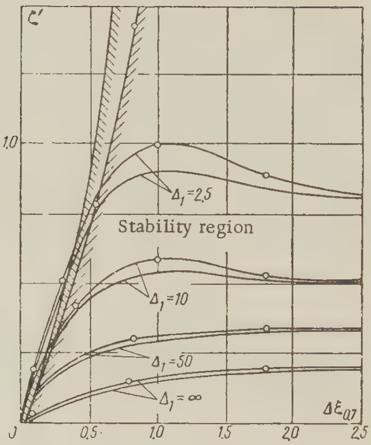


Fig. 9. Dependence of relative noise bandwidth of automatic phase control with proportional-integrating low-pass filter on the IF amplifier bandwidth for different values of Δ_1 for $\eta = 0.2$. Solid curves — n = 1; circles — n = 10.

These relationships, calculated for one and for ten amplifier stages (n = 1 and n = 10) at two values of the low-pass filter constant, $\Delta_1 = 5$ and $\Delta_1 = 10$, are plotted in Fig. 8.

From an examination of this figure it is seen that as the bandwidth of the IF amplifier is narrowed down beyond the limits of the bandwidth of the closed-loop automatic phase control system, the noise band increases to a definite maximum value, which lies near the cutoff frequency of the closed-loop characteristic. Upon further reduction in the bandwidth of the IF amplifier, the noise band Π_n decreases sharply. However, by considerably reducing the IF amplifier bandwidth, it is impossible to improve the filtering properties of the system, since the limit of the stability region passes near the peak of the curve $\Pi_n(\Delta\xi_{0.7})$, and reduction of the bandwidth leads to violation of inequality (6) and to excitation of the system (the stability region I is plotted for n = 1, while stability region II is for n = 10). The situation is different if we use a proportional-integrating filter for the low-pass filter. In this case, by virtue of (11), (15) and (19) the expression for the noise band can be represented in the form

$$\Pi^{*}_{in} = \frac{1}{\tau_0}\int\limits_0^\infty \frac{(1+\Delta_2^2\xi^2)\,d\xi}{\left[1+(1+\Delta_2^2)\,\xi^2+\Delta_1^2\,\xi^4+(\Delta_2-\Delta_1)\cos\xi\,\frac{\alpha_1(n)}{\Delta\xi_{0,7}}\times\right.} \rightarrow$$

$$\rightarrow \frac{(1+\Delta_2^2\xi^2)\,d\xi}{\left.\times\, 2\xi^2-2\xi\,(1-\Delta_1\Delta_2\xi^2)\sin\xi\,\frac{\alpha_1(n)}{\Delta\xi_{0,7}}\right]\left[1+\frac{2\xi}{\Delta\xi_{0,7}}\,\frac{1}{\psi(n)}\right]^2} \qquad (22)$$

By numerically integrating the resultant expression (21), we have plotted the dependence of Π'_n on $\Delta\xi_{0.7}$ for different values of Δ_1 ($\Delta_1 = 2.5$, 10, 50, and ∞) and different values of n (n = 1, 10) with $\eta = 0.2$ (Fig. 9).

As can be seen from this plot, the character of variation of Π^*_n with varying $\Delta\xi_{0.7}$ is essentially analogous to the preceding case. At small values of $\Delta\xi_{0.7}$, however, the drooping portion of the characteristic $\Pi'_n(\Delta\xi_{0.7})$ lies within the limits of the stability region of the system (at sufficiently large values of Δ_1). By virtue of this, we can improve in this case the filtering properties of the automatic phase control system by considerably narrowing down the IF amplifier bandwidth, compared with the closed-loop bandwidth. Apparently, an analogous analysis can be made for any other IF amplifier circuit.

3. QUALITY OF REGULATION OF THE AUTOMATIC PHASE CONTROL SYSTEM

As shown above, the accuracy of phase control in the steady state can be sufficiently high if the system parameters are suitably chosen. It is, therefore, of interest to examine the behavior of the system in the dynamic mode, for in this case the decisive error becomes the transient error. This error can be characterized in our analysis by the two following quantities: the excess of the maximum value of detuning over the steady state value, the so-called over-regulation σ, and the time interval after which the detuning reaches the steady-state value for the first time, the settling time t_s.

a) Estimate of the value of over-regulation. To estimate the value of the over-regulation we used the well known Solodovnikov formula [9]

$$\sigma \leqslant 100\,(1.18\,\mu - 1)\,\%, \tag{23}$$

where $\mu = \max |W(j\omega)| \,/\, |W(0)|$ is the index of "oscillatoriness".

By virtue of (2), the expressions for the frequency characteristics of the phase control system has the form

$$|W(j\omega)| = \frac{1}{\sqrt{1 - 2\xi \sin \Delta_3\xi + (1 - 2\Delta_1 \cos \Delta_3\xi)\,\xi^2 + \Delta_1^2\,\xi^4}} \tag{24}$$

for a system with a low-pass integrating filter, and

$$|W^*(j\xi)| = \sqrt{\frac{1 + \Delta_2^2\,\xi^2}{1 + (1 + \Delta_3^2)\,\xi^2 + \Delta_1^2\,\xi^4 + 2\xi^2\,(\Delta_2 - \Delta_1)\cos \Delta_3\xi - 2\xi\,(1 + \Delta_1\Delta_3\xi^2)\sin \Delta_3\,\xi}} \tag{25}$$

for a system with a proportional-integrating low-pass filter. When $\Delta_2 = 0$ the expression (25) reduces to (24), and in the absence of delay ($\Delta_3 = 0$) expressions (24) and (23) become the same as in [3].

Although generally speaking the expression for the quantity μ, contained in (23), can be determined analytically, it is simpler to evaluate it from the frequency characteristics plotted on the basis of (24) and (25) for each specific combination of the parameters.* Fig. 10 shows the dependence of the over regulation σ in a system with an integrating and proportional-integrating low-pass filter on the delay and on the low-pass filter parameters (Δ_1 and η).

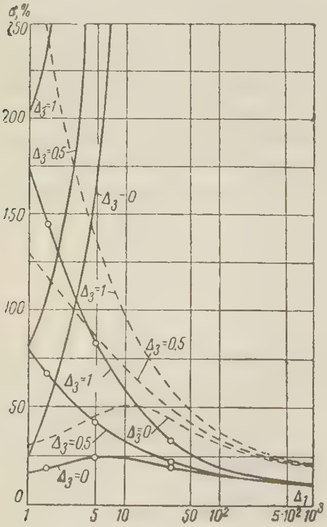


Fig. 10. Dependence of the over regulation on the delay time and the parameters of the low-pass filter Δ_1 and η. Solid curve — integrating filter; dashed curves — proportional-integrating filter ($\eta = 0.2$); circles — proportional-integrating filter ($\eta = 0.4$).

From an examination of the curves shown in Fig. 10, we can note that in a phase control system with integrating low-pass filter, even at relatively small values of the filter constant Δ_1, the over regulation reaches a considerable magnitude. When Δ_1 increases and the delay time Δ_3 increases, the over regulation in the system increases sharply and the system becomes practically unusable. If a proportional-integrating low-pass filter is used in the system, the opposite

---

*The frequency characteristics were already plotted above, in the determination of Π_n and Π^*_n.

is observed. In this case for any specified value $\tau_3 < \tau_{cr}$ we can make the over regulation, generally speaking, as small as desired, by choosing for this purpose sufficiently large values of Δ_1 and Δ_2 and a sufficiently small η.

b) Estimate of the transient time. In the general case, the transient time must apparently be estimated from the plotted regulation curve, by virtue of the relative complexity of the expressions for the transfer functions. However, in an oscillating transient we can obtain sufficiently simple approximate expressions for the transient time. For this purpose we make use of the following expression [9] for the dimensionless settling time $\Delta_s = t_s/\tau$:

$$\Delta_s = \frac{\pi}{\xi_1}, \tag{26}$$

where ξ_1 is the root of the equation $|\arg W(j\xi)| = \pi/2$. In the general case the value of ξ_1 is determined from the phase characteristics of the automatic control system. In our case, however, the problem can also be solved analytically.

For a system with an integrating low-pass filter, the equation for ξ_1 assumes by virtue of (2) the form

$$\Delta\xi_1^2 \cos \Delta_3\xi + \xi_1 \sin \Delta_3\xi_1 - 1 = 0,$$

from which we obtain for small Δ_3, with account of (25),

$$\Delta_y \simeq \pi\sqrt{\Delta_1}\left(1 + \frac{\Delta_3}{2\Delta_1}\right). \tag{27}$$

It is seen from the last expression that the presence of delay in the system increases the settling time, and that this effect decreases with increasing Δ_1. At sufficiently large values of $\Delta_1 (\Delta_1 \gg \Delta_3)$, Δ_s becomes directly proportional to $\sqrt{\Delta_1}$. For a system with proportional-integrating low-pass filter, the equation for ξ_1 has the form

$$1 + \eta^2\Delta_1^2\xi_1^4 + \Delta_1(\eta - 1)\xi_1^2 \cos \Delta_3\xi_1 - \xi_1(1 + \Delta_1^2\eta\xi_1^2) \sin \Delta_3\xi_1 = 0. \tag{28}$$

This equation has a simple solution only in the absence of delay

$$\Delta_s = \pi\sqrt{\Delta_1}\sqrt{1 - \eta(1 + \eta\Delta_1)}. \tag{29}$$

In the general case, on the other hand, the solution must be found graphically.

Fig. 11 shows the dependence of the settling time Δ_s on the filter constant Δ_1 at fixed values of η ($\eta = 0.1$ and 0.2) for a system with a proportional-integrating and with an integrating low-pass filter. From the curves of Fig. 11 it is seen that the settling time increases in this case with increasing delay time and that the settling time increases more sharply with increasing Δ_1. It must be noted that these conclusions, as well as expressions (26) and (28), are valid only when the transient is oscillatory. For a general investigation of the character of the transient it is necessary to plot the regulation curves.

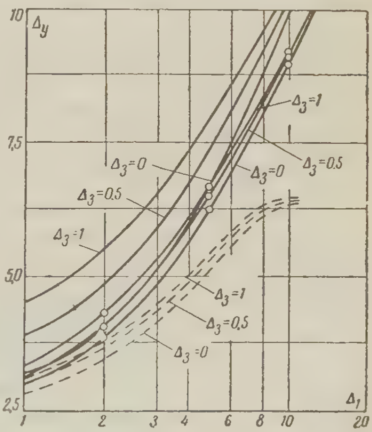


Fig. 11. Dependence of the settling time on the parameter Δ_1 and on the delay time. Solid curves — circuit with integrating filter; circles — circuit with proportional-integrating filter ($\eta = 0.1$); dashed curves — circuit with proportional integrating filter ($\eta = 0.2$).

c) Regulation curves. To plot the regulation curves of an automatic phase control system with delay, we first constructed the semi-logarithmic real characteristics of the system. After piecewise

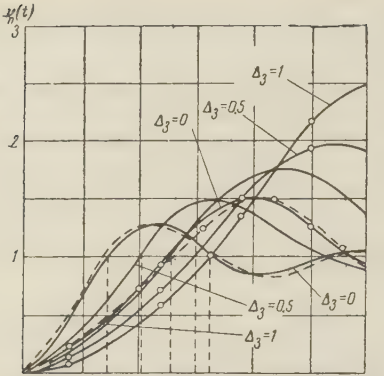


Figure 12

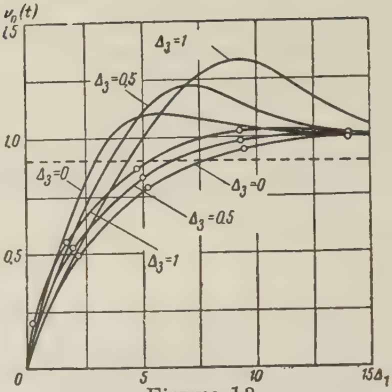


Figure 13

Fig. 12. Transfer function of automatic phase control system with integrating low-pass filter for different values of delay time. Solid curves — $\Delta_1 = 2.5$; circles — $\Delta_1 = 5.0$.

Fig. 13. Transfer function in the case of a system with proportional-integrating low-pass filter with $\eta = 0.4$. Solid curves — $\Delta_1 = 2.5$; circles — $\Delta_1 \to \infty$.

linear approximation we plotted by the well known method the transient characteristics of the system with integrating low-pass filter (Fig. 12) and with proportional-integrating filter (Fig. 13) for different values of the constant Δ_3 and of the dimensionless running time $\Delta = t/\tau$.

From an examination of the curves we can draw the following conclusions.

In a system of automatic phase control with integrating low-pass filter, the transient is oscillatory. The settling time and the over regulation increase with increasing delay time and with increasing Δ_1. In the case when a proportional-integrating filter is used the variation of t_s and σ with Δ_3 and Δ_1 is analogous to the variation in the preceding case if the process is oscillatory. As Δ_1 is increased with fixed η, the transient becomes aperiodic and Δ_s, although increasing with increasing Δ_1, nevertheless decreases as the delay time Δ_3 is increased. In other words, in this case the presence of delay in the channel increases the speed of the automatic phase control system.

4. EXPERIMENTAL RESULTS

An experimental verification of the foregoing conclusions was made with a model of an automatic phase control system for the frequency of a klystron generator (Fig. 14) operating in the $\lambda = 15$ cm band.

The reference IF signal used was produced by a crystal-controlled oscillator at a frequency $f_0 = 26.5$ Mc; the 78th harmonic of these oscillations, obtained by multiplying the frequency with vacuum tube and diode stages, served as the microwave reference signal with a power on the order of $P_s = 1$ microwatt. This sytem ensured stable tuning of the klystron oscillator at a holding band of 60 Mc, a locking band $\Delta\omega 3 = 5$ Mc, and a transient time $t_s = 10$ microseconds.

To estimate the correctness of the conclusions obtained above, we used the model to plot the frequency characteristics of the closed-loop automatic phase control system with proportional-integrating low-pass filter, for two different values of the ratio η, and for several values of the time constant Δ_1. These characteristics are shown in Figs. 15 and 16, where the calculated frequency characteristics, calculated for the same numerical values of the parameters on the basis of formula (25), are shown for comparison. The discrepancy between the numerical values of the theoretical and experimental curves does not exceed 15%, and for the lower value of η this discrepancy decreases to 6%. Confirmation is also obtained for the qualitative conclusions of the foregoing study of the character of the influence of the parameters Δ_1 and Δ_2 on the frequency characteristics of the closed-loop automatic phase control

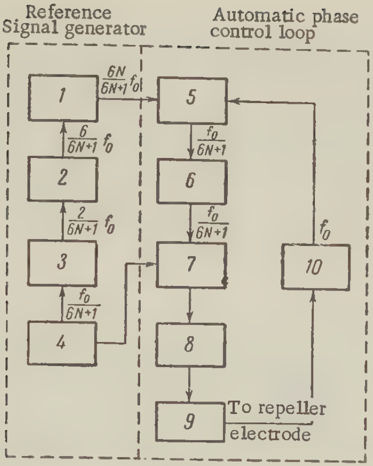


Fig. 14

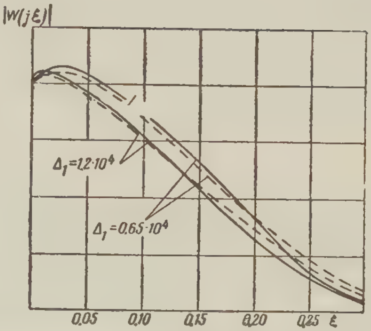


Fig. 15

Fig. 14. Block diagram of breadboard of automatic phase control system for a klystron generator. 1 — semiconductor diode; 2 — frequency tripler; 3 — frequency doubler; 4 — crystal stabilized oscillator; 5 — mixer; 6 — IF amplifier; 7 — phase detector; 8 — low-pass filter; 9 — IF amplifier; 10 — klystron oscillator.

Fig. 15. Frequency characteristics of closed-loop automatic phase control system with proportional-integrating low-pass filter (η = = 0.13; $\Delta_3 = 6$). Solid curves — theoretical values; dashes — experimental values

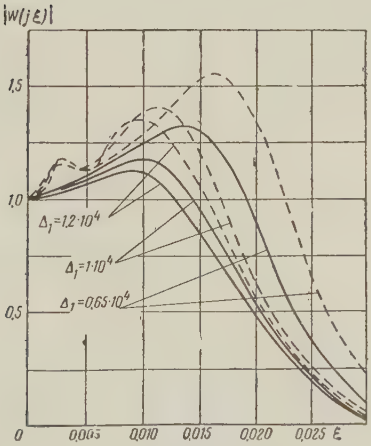


Fig. 16

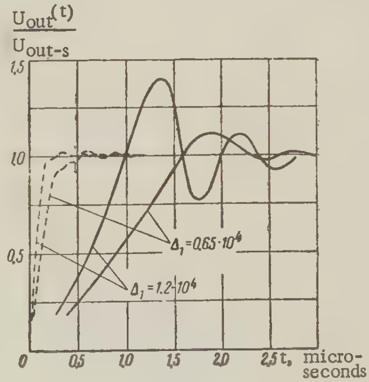


Fig. 17

Fig. 16. Frequency characteristics of closed-loop automatic phase control system with proportional-integrating low-pass filter (η = = 0.006; $\Delta_3 = 6$). Solid curves — theoretical values; dashes — experimental values.

Fig. 17. Transient characteristics of automatic phase control system (experimental); solid curves — $\eta = 0.006$; dashed curve — $\eta = 0.13$.

system, and consequently on the filtering properties of the system. For a qualitative verification of the conclusions regarding the effect of the parameters Δ_1 and Δ_2 on the quality of regulation of the system, the experimental model was used to plot the transient characteristics of the closed-loop automatic phase control system (Fig. 17) for two different values of the parameters Δ_1 and Δ_2. It follows from Fig. 17 that the over regulation in the system decreases with decreasing filter constant Δ_1 (at sufficiently large absolute values of the latter) and with increasing constant Δ_2. An increase in the constant Δ_2 also leads to an increase in the speed of the system and thus to an improvement in the regulation quality of the system as a whole. These results are in good agreement with the theoretical conclusions of Section 3 of the present paper.

REFERENCES

1. M.R. Kaplanov, V.L. Levin. Automatic Frequency Control. GEI, 1956.
2. I.L. Bershteyn V.L. Sibiryakov. Automatic phase control of the frequency of centimeter-wave generators. Radiotekhnika i elektronika, 1958, 3, 2, 290.
3. M.V. Kapranov. Automatic phase control of frequency. Candidates Dissertation, 1957.
4. M.V. Kapranov. On the locking band of automatic phase control of frequency. Radiotekhnika, 1956, 11, 12.
5. L.S. Tyufyakin. Automatic phase control system. Scientific Notes of the Higher Schools—Radio Engineering, 1958, 1, 1.
6. T.S. George. Analysis of synchronizing systems for of interlaced color television. Proc. IRE, 1951, 39, 12.
7. W.J. Gruen. Theory of AFC synchronization. Proc. IRE, 1953, 41, 8.
8. V.A. Shpolyanskiy. Investigation of automatic phase control of the frequency of generators for the centimeter band. Diploma Project, Moscow Power Institute, 1959.
9. V.V. Solodovnikov. Frequency method of analysis of quality of automatic control system. Avtomatika i telemekhanika. 1952, 3.
10. L.A. Birger. Stability of an automatic phase control circuit for a klystron. Radiotekhnika. 1961, 16, 3.
11. M.A. Volin. Intermediate Frequency Amplifiers. Soviet Radio Press, 1950.

Received by editor 1 March 1960

ANALYSIS OF ANTENNAS WITH PLANE APERTURES

B.M. Minkovich, Yu.I. Davidchevskiy

A method is developed for the synthesis of antennas with plane apertures (calculation of the amplitude-phase or current distribution and of the shape of the aperture), based on the use of the results of synthesis of a linear antenna.

General solutions are given for the case of separating distributions when the directivity pattern is specified in one or in two principal planes. One structural solution is indicated for the case of non-separating distributions.

INTRODUCTION

In many applications, antennas must have radiation characteristics that satisfy specific requirements. Such characteristics are the maximum directivity coefficient, the form of the entire directivity pattern or of its principal lobe necessary for optimum relationship between the different regions, or the parameters of the directivity pattern.

The required radiation characteristics of an antenna is determined by various methods, including the aperture method, which is the simplest in most cases. In the aperture method, a system of monopolarized dipoles, which produces the interference pattern in space, is placed in a plane aperture of definite form. The determination of the directivity pattern of an antenna from the specified amplitude-phase distribution of the radiation sources and specified form of the aperture is the direct problem in the synthesis of the directivity pattern.

The reverse problem in the directivity-pattern synthesis, or the direct problem in aperture synthesis, is the determination of the amplitude-phase distribution and the form of the antenna aperture from specified characteristics or parameters of the directivity pattern.

The synthesis of antennas with plane apertures comprises two problems, the first being the determination of the amplitude-phase distribution for a specified antenna aperture, and the second the determination of the form of the antenna aperture for a known amplitude-phase distribution.

The first synthesis problem has been most fully developed for a linear antenna, which can be represented as a rectangular aperture of unit width with amplitude and phase distributions that are constant over the width.

Less developed is the first problem for antennas having apertures of different shapes. Many problems, including even that for a round-aperture antenna, have been solved only recently.

Individual solutions of the second synthesis problem are now beginning to appear. Thus, the choice of aperture shapes for the secondary radiator and the radiator of periscopic antennas is considered in [1, 2], and the form of the antenna-length modulation curve for four-dimensional radiators is considered in [3]. The general approach to the solution of such problems, however, has not been sufficiently developed. We propose below a method for the synthesis of antennas with plane apertures, wherein the general problem of synthesis can be solved by simultaneous choice of the amplitude-phase distribution and of the shape of antenna aperture.

1. FUNDAMENTAL EQUATIONS

In the aperture method, the directivity pattern of an antenna in the far zone is calculated from the expression [4]

$$F(\varphi, \theta) = \int_{\dot{D}} \dot{A}(\xi, \eta)\, e^{j(u_1 \xi \cos\varphi + u_2 \eta \sin\varphi)}\, d\xi\, d\eta, \tag{1}$$

where $A(\xi, \eta)$ is the amplitude-phase distribution with linear polarization; ξ, η — relative Cartesian coordinates of the aperture;

$$u_1 = \frac{\pi l_1}{\lambda} \sin\theta, \qquad (2)$$
$$u_2 = \frac{\pi l_2}{\lambda} \sin\theta;$$

where l_1 and l_2 are the dimensions of the real aperture along the axes ξ and η, respectively.

In the principal planes ($\varphi = 0$ and $\varphi = \pi/2$), expression (1) assumes the form:

$$F(0, \theta) = \int_{-1}^{1} \left[\int_{b_2(\xi)}^{b_1(\xi)} \dot{A}(\xi, \eta)\, d\eta \right] e^{ju_1\xi}\, d\xi, \qquad (3)$$

$$F\left(\frac{\pi}{2},\ \theta\right) = \int_{-1}^{1} \left[\int_{a_2(\eta)}^{a_1(\eta)} \dot{A}(\xi, \eta)\, d\eta \right] e^{ju_2\eta}\, d\eta, \qquad (4)$$

where $\eta = b(\xi)$ and $\xi = a(\eta)$ are the equations of the envelope of the aperture.

If it is recalled that the directivity pattern of a linear antenna has the form

$$\int_{-1}^{1} \dot{J}_1(\xi) e^{ju\xi}\, d\xi, \qquad (5)$$

then a comparison of (3) and (4) with (5) shows that the expressions in the square brackets are equivalent to the amplitude-phase distribution of certain linear antennas. This is a known fact [1]; from the point of view of the directivity pattern in one of the planes, any plane antenna can be replaced by an equivalent linear antenna, whose effective amplitude-phase distribution is determined by the form of the plane aperture. An antenna with plane aperture and a linear antenna have identical diagrams in the "cross-sectional plane" of a three-dimensional directivity pattern.

Starting from (3), (4), and (5), we get

$$\int_{b_2(\xi)}^{b_1(\xi)} \dot{A}(\xi, \eta)\, d\eta = k_1 \dot{J}_1(\xi), \qquad (6)$$

$$\int_{a_2(\eta)}^{a_1(\eta)} \dot{A}(\xi, \eta)\, d\xi = k_2 \dot{J}_2(\eta), \qquad (7)$$

where $\dot{J}_1(\xi)$ and $\dot{J}_2(\eta)$ are the amplitude-phase distributions of the equivalent linear antennas; k_1 and k_2 are certain constants.

Eqs. (6) and (7) enable us to calculate the effective distributions of antennas with plane aperture in two orthogonal (principal) planes. However, if it is assumed that $\dot{J}_1(\xi)$ and $\dot{J}_2(\eta)$ are known, then (6) and (7) can be regarded as a system of an equation with which to synthesize an antenna with plane aperture from known directivity patterns in the principal planes.

Methods for the synthesis of the amplitude-phase distribution of a linear antenna from a known directivity pattern have by now been sufficiently well developed [5, 6, 7]. Consequently, from the pattern of a plane-aperture antenna in the planes $\varphi = 0$ and $\varphi = \pi/2$ we can synthesize $\dot{J}_1(\xi)$ and $\dot{J}_2(\eta)$. Now in turn, the system (6) and (7) can serve for the synthesis of the entire plane aperture of the antenna.

In such a synthesis, the solution of the problem is frequently not unique: several amplitude-phase distributions and several antenna-aperture shapes are possible and produce the required directivity pattern. In most cases the synthesis is incomplete, since it is based not on the complete diagram, but only on two of its cross sections. However, in view of the fact that the synthesis of antennas with plane aperture has been but little developed, an antenna-aperture synthesis method based on two directivity-pattern sections can be of certain value, all the more since the antenna directivity pattern is frequently deduced from particular diagrams in two characteristic sections.

We shall henceforth assume $\dot{J}_1(\xi)$ and $\dot{J}_2(\eta)$ known.

Let us investigate Eqs. (6) and (7), so as to ascertain whether a plane-aperture antenna can be synthesized, i.e., whether $\dot{A}(\xi, \eta)$ and $b(\xi)$ can be determined from known $\dot{J}_1(\xi)$ and $\dot{J}_2(\eta)$.

The proposed synthesis method enables us to relate plane-aperture antennas with linear antennas and make use of the powerful formalism for the synthesis of linear antennas in the synthesis of plane-aperture antennas.

2. TYPES OF AMPLITUDE-PHASE DISTRIBUTION AND OF APERTURE SHAPES

Eqs. (6) and (7) call for a correct specification of the functions $\dot{J}_1(\xi)$, $\dot{J}_2(\eta)$, $b(\xi)$, $a(\eta)$. Actually, if the envelope of the aperture profile decreases to zero ($b_1(\xi)$ and $b_2(\xi)$ equal to zero when $\xi = 1$), then $\dot{J}_1(\xi)$ should also drop to zero on the edge of the aperture; otherwise $\dot{A}(\xi, \eta)$ would become infinite. Therefore, $\dot{J}_1(\xi)$ and $\dot{J}_2(\eta)$ must be specified with allowance for the shape of the aperture. Thus, for example, in an antenna with round aperture, it is actually impossible to obtain an effective distribution with "pedestal" when $\xi = 1$. An effective amplitude-phase distribution with pedestal can be realized in antennas whose aperture has a bevel on the edge, i.e., it does not drop to zero, or with the aid of an amplitude-phase distribution that becomes infinite on the edges of the aperture.

Thus, for example in an antenna with round aperture, it is actually impossible to obtain an effective distribution with "pedestal" when $\xi = 1$. An effective amplitude-phase distribution with pedestal can be realized in antennas whose aperture has a bevel on the edge, i.e., it does not drop to zero, or with the aid of an amplitude-phase distribution that becomes infinite on the edges of the aperture.

Not only the amplitude and phase distributions, but also the shape of the antenna aperture greatly influence the directivity pattern. The shape of the antenna aperture enters into (6) and (7) as a full fledged term, and therefore synthesis of an antenna with plane aperture includes in general both the shape of the aperture and the distribution of the amplitudes and phases.

We shall henceforth distinguish between two classes of amplitude and phase distributions, separating and non-separating. We shall call a separating distribution one of the type

$$\dot{A}(\xi, \eta) = \dot{A}_1(\xi)\, \dot{A}_2(\eta).$$

Although amplitude distributions can be symmetrical and asymmetrical, in phase and out of phase, we shall pay principal attention to symmetrical in-phase distributions.

We shall consider two types of problems.

In the first, the directivity pattern of the plane antenna is specified only in one plane, and consequently only one of the equations, either (6) or (7), remains. Many data on problems of this type are contained in [1].

The second type is one in which the plane-antenna directivity pattern is specified in two planes, and consequently both equations (6) and (7) must be solved simultaneously.

3. SEPARATING DISTRIBUTIONS

1) First type of problem. The aperture ($b(\xi)$) is symmetrical with respect to the ξ axis.

a) In-phase symmetrical distributions. The problem reduces to solving the equation

$$A_1(\xi) \int_0^{b(\xi)} A_2(\eta)\, d\eta = J_1(\xi). \tag{8}$$

Since there is only one equation and three unknown functions $A_1(\xi)$, $A_2(\eta)$, and $b(\xi)$, two of these must be specified in order for the third to be found. This leads to the following three equations for determining one function when the other two are specified:

$$A_1(\xi) = \frac{J_1(\xi)}{\int\limits_0^{b(\xi)} A_2(\eta)\,d\eta}, \tag{9}$$

$$A_2(\eta) = \left\{\left[\frac{J_1(\xi)}{A_1(\xi)}\right]' \frac{1}{b'(\xi)}\right\}_{\xi=a(\eta)}, \tag{10}$$

$$\Phi_2[b(\xi)] = \frac{J_1(\xi)}{A_1(\xi)}, \tag{11}$$

where $\Phi_2(\eta) = \int\limits_0^{\eta} A_2(\eta)\,d\eta$. The prime denotes differentiation with respect to ξ.

Expressions (9), (10) and (11) enable us to determine, from the known amplitude distribution of the linear antenna, the amplitude distribution of the plane antenna necessary to ensure the required directivity in one of the planes. The fact that the problem has manifold solutions offers the advantage that we can choose for a given specific case the most expedient from among several different distributions and aperture shapes.

We note that Eq. (8) has a simple physical interpretation: the effective distribution $J_1(\xi)$ is equal to the distribution of a plane antenna along the direction of the ξ axis, multiplied at each point by the corresponding quantity, proportional to the effective moment of the cross section of the antenna in the direction of the η axis [8].

b) Asymmetrical in-phase distributions. Eq. (8) shows that if $J_1(\xi)$ is chosen symmetrical and in phase, then independently of the choice of $A_1(\xi)$, if $A_2(\eta)$ and $b(\xi)$ are symmetrical with respect to the ξ axis, the effective distribution $J_2(\eta)$ will also be symmetrical, and the phase center of a plane-aperture antenna, with respect to the two principal planes, is at the origin. The calculation of the amplitude-phase distribution is based on formulas (9), (10), and (11).

c) Symmetrical out-of-phase distributions. For an odd phase distribution $\varphi_2(\eta)$, the principal equation has the form

$$A_1(\xi)e^{j\psi_1(\xi)} \int\limits_0^{b(\xi)} A_2(\eta)\cos\,\psi_2(\eta)\,d\eta = J_1(\xi)\,e^{j\varphi_1(\xi)}. \tag{12}$$

Consequently, $\psi_1(\xi) = \varphi_1(\xi)$; $\psi_2(\eta)$ influences the effective distribution of the amplitudes. To solve (12) we must specify not two but three functions and determine the fourth one by means of formulas similar to (9), (10), and (11). Analogous but more cumbersome formulas can be used to solve the problem for other out-of-phase distributions.

In the case of asymmetrical out of phase distributions, the principal equation actually becomes two, for negative and positive ξ separately. The mathematical problem becomes more complicated, but can be solved in each specific case.

2) Second type of problem. The aperture is symmetrical with respect to the ξ and η axes.

a) In-phase symmetrical distributions. The problem reduced to solving simultaneously the two equations

$$A_1(\xi) \int\limits_0^{b(\xi)} A_2(\eta)\,d\eta = k_1 J_1(\xi), \tag{13}$$

$$A_2(\eta) \int\limits_0^{a(\eta)} A_1(\xi)\,d\xi = k_2 J_2(\eta), \tag{14}$$

where $b(\xi)$ and $a(\eta)$ are mutually reciprocal functions, i.e.,

$$b(\xi) = a^{-1}(\xi).$$

Eqs. (13) and (14) can be solved in general form. To solve them simultaneously we must specify one of the functions $A_1(\xi)$, $A_2(\eta)$, or $b(\xi)$.

$A_1(\xi)$ specified. It is required to find $b(\xi)$ and $A_2(\eta)$. As a result of simple transformations we obtain the following equation for $y = b(\xi)$:

$$k_2 \int J_2(y)\, dy = k_1 \int \Phi_1(\xi) \left[\frac{J_1(\xi)}{A_1(\xi)} \right]' d\xi + C, \tag{15}$$

where $\Phi_1(\xi) = \int_0^{\xi} A_1(\xi)\, d\xi$; C is a constant.

From (12) we have

$$A_2(\eta) = \frac{k_2 J_2(\eta)}{\Phi_1[a(\eta)]}. \tag{16}$$

b(ξ) specified. It is required to find $A_1(\xi)$ and $A_2(\eta)$. By determining $A_2(\eta)$ from (14) and substituting in (13), subsequent differentiation reduces the determination of $A_1(\xi)$ to the solution of the following differential equation

$$ZZ'' + k_1 \frac{J_1(\xi)}{B(\xi)} Z'^2 - \frac{B'(\xi)}{B(\xi)} ZZ' = 0, \tag{17}$$

where

$$Z = \frac{J_1(\xi)}{A_1(\xi)}; \tag{18}$$

$$B(\xi) = k_2 J_2[b(\xi)]\, b'(\xi). \tag{19}$$

Eq. (17) can be transformed to the Bernoulli type and can be solved in general form:

$$A_1(\xi) = C_2 J_1(\xi) \exp \left\{ - \int \frac{B(\xi)\, d\xi}{k_1 \int J_1(\xi)\, d\xi + k_2 \int J_2[b(\xi)]\, db(\xi) + C_1} \right\}, \tag{20}$$

where C_1 and C_2 are constants; $A_2(\eta)$ is determined from (16).

We note that Eqs. (6) and (7) can also be used to calculate some asymmetrical out of phase distributions. For example, for odd functions $\varphi_1(\xi)$ and $\varphi_2(\eta)$, it is sufficient to replace in the exponential factor of (19) and (20) $J_1(\xi)$ by $J_1(\xi) \cos \varphi_1(1)$ and $J_2(\eta)$ by $J_2(\eta) \cos \varphi_2(\eta)$, where $\varphi_1(\xi)$ and $\varphi_2(\eta)$ are the phase distributions.

The foregoing results show that the case of separating distributions offers in principle no difficulties in the synthesis of plane-aperture antennas from the directivity pattern in two or one principal section, although mathematical difficulties may arise in the evaluation of the integrals in (20).

4. NONSEPARATION DISTRIBUTIONS

In the case of separating distributions, the structure of the distributions of the amplitude is rather simple: the intersection of the amplitude distribution with one of the coordinate planes actually determines the form of the intersections with the planes parallel to the coordinate plane. In the case of non-separating distributions, the intersection between the amplitude distribution and one plane gives very little information on the form of the intersections with other planes. Consequently it is hardly possible to solve the problems for non-separating distributions in general form, and the situation is much more complicated.

1) First type of problems.

a) Symmetrical in-phase distributions. It is required to solve the equation

$$\int_0^{b(\xi)} A(\xi, \eta)\, d\eta = J_1(\xi). \tag{21}$$

1) If $A(\xi, \eta)$ is specified, we obtain after integration in (21) the equation for the determination of $b(\xi)$.

2) If $b(\xi)$ is specified, Eq. (21) cannot be solved in general form and has an infinite number of solutions.

We can recommend the following solutions of Eq. (21):

$$\text{a)}\ A(\xi, \eta) = \sum_{k=0}^{N} \frac{a_k}{F_k[b(\xi)]} f_k(\eta) J_1(\xi), \tag{22}$$

where $f_k(\eta)$ is an arbitrarily readily integrable function;

$$F_k(\xi) = \int_0^{\xi} f_k(\eta)\, d\eta;$$

$\sum_{k=0}^{N} a_k = 1$ for chosen N.

$$\text{б)}\ A(r_1) = -\frac{dJ_1[a(1-r_1)]}{dr_1}, \tag{23}$$

where $r_1 = 1 + \eta - b(\xi)$.

$$\text{в)}\ A(r_2) = \frac{2}{\pi}\left[\frac{J_1[a(0)]}{\sqrt{1-r_2^2}} - \int_{r_2}^{1} \frac{dJ_1[a(\sqrt{1-Z^2})]}{\sqrt{Z^2-r_2^2}}\right], \tag{24}$$

where $r_2 = \sqrt{1 + \eta^2 - [b(\xi)]^2}$; $a(\eta)$ is a function reciprocal to $b(\xi)$.

Of course, expressions (22), (23), and (24) do not exhaust the entire set of solutions.

Even for the case of in-phase nonseparating distributions, it is difficult to make some general recommendations for a simultaneous solution of Eqs. (6) and (7). However, if we seek the solution in the form (22), then $f_k(\eta)$ should satisfy an equation similar to (17).

5. "ARTIFICIAL" BEAM ROCKING

If the directivity pattern of an antenna is calculated in one plane, the question arises of determining the complete directivity pattern.

A useful method here is that of linear artificial rocking of the beam, consisting in the following.

We assume that for a specified diagram in one plane $\varphi = 0$ we solve Eq. (8), i.e., we know $b(\xi)$ and $A(\xi, \eta)$. We assume that in the plane $\varphi = \pi/2$, the function $A(\xi, \eta)$ now acquires a linear phase shift. Then the new effective distribution is

$$\int_0^{b(\xi)} A(\xi, \eta) \cos \alpha\eta\, d\eta = J_{1\alpha}(\xi). \tag{25}$$

Since the linear phase shift causes a shift of the principal beam in the entire directivity pattern in the generalized-coordinate system, Eq. (25) is the effective distribution of a linear antenna, the directivity pattern of which coincides with the directivity pattern of an antenna with a plane aperture in the section $u_2 = \alpha$. Thus, by taking different α, we can investigate the directivity pattern, as it were, over a series of sections.

CONCLUSION

The synthesis method considered for antennas with plane aperture, on the basis of one or two sections of the directivity pattern, enables us to use the natural connection between an aperture of arbitrary shape and linear antennas. Many synthesis problems, formulated and solved for linear antennas, can be solved also for surface apertures. Thus, for example, the problem of optimum or quasi-optimum diagrams can be solved either by producing in a specified aperture a definite (optimal or quasi-optimal) amplitude-phase distribution, or by choosing an aperture with optimal or quasi-optimal shape for a specified distribution [8].

An illustrative example of the use of the method is the synthesis of amplitude distribution in a round aperture [9].

Consequently, in certain problems, the method developed here can prove useful and effective.

REFERENCES

1. G. F. Koch. Flächenstrahler mit kleinen Nebenmaxima, Fernmeldetechn. Z., 1954, 7, 10, 498.
2. A.M. Pokras, Choice of form of radiator and reradiator aperture. Elektrosvyaz', 1959, 3, 41.
3. H.E. Shanks and R.W. Bickmore. Four-dimensional radiators for electromagnetic waves, Canad. J. Phys. 37, 263 (1959).
4. Microwave Antennas, translated from the English. (Probably MIT Radiation Laboratory Series, v. 12).
5. N.V. Kovalenko. Determination of current of linear antenna for a specified directivity pattern, NDVSh (Scientific Report of the Universities) — Radio Engineering and Electronics, 1959, 2, 94.
6. L.B. Tartakovskiy, Synthesis of linear radiator and its analogues in the problem of broadband matching. Radiotekhnika i elektronika, 1958, 3, 12, 1463.
7. G.H. Brown. Pattern synthesis-simplified methods of array design to obtain a desired directive pattern. RCA Rev., 1959, 20, 3, 398.
8. Yu.I. Davidchevskiy, B.M. Minkovich, V.A. Kalinin, Concerning antennas with quasi-optimal apertures. Radiotekhnika i elektronika, 1960, 5, 9, 1545.
9. B.M. Minkovich, Yu.I. Davidchevskii. On the synthesis of antennas with round apertures, Radiotekhnika i elektronika. 1961, 6, 8, 1395.

Received by editor 5 August 1960

TRANSVERSE DIFFUSION IN DIFFRACTION BY AN IMPEDANCE CYLINDER OF LARGE RADIUS.

II. ASYMPTOTIC LAWS OF DIFFRACTION IN POLAR COORDINATES

L. A. Vaynshteyn, G. D. Malyuzhinets

It is shown that the general solution of the diffraction problem, obtained in [1] by using a parabolic equation in ray coordinates, can be derived also from an exact solution of the wave equation by using asymptotic formulas for the Hankel functions. We discuss incidentally problems (such as the reflection formula and its refinement, or the surface wave) an analysis of which with the aid of the parabolic equation and the ray coordinates entails certain difficulties.

INTRODUCTION

In Part I of our paper [1] we used the parabolic equation in ray coordinates to obtain a general solution of the two-dimensional diffraction problem for an impedance cylinder, whose radius is much greater than the wavelength. By way of supplement to this result, we shall show below that this solution can be obtained from an exact solution of the wave equation, by using known asymptotic formulas for the Hankel functions.

The exact solution enables us also to investigate problems connected with the "transverse diffusion" in the illuminated region, when the reflection coefficient vanishes for some angle of incidence (Section 2), and also problems connected with the excitation and propagation of surface waves (Section 3). It is so far impossible to tackle these problems with the aid of the parabolic equation and the ray coordinates.

1. THE GREEN'S FUNCTION GAMMA FOR THE IMPEDANCE CYLINDER

The problem is stated in Section 1 of reference [1]. It reduces to constructing a function $\Gamma(r, \varphi, r')$ in a plane of many sheets. Then the Green's function in the physical plane is obtained by summing the values of the function Γ over all the sheets (see formula (8) of [1]). The function Γ can be readily obtained in the form of a series

$$\Gamma(r, \varphi, r') = -\frac{4\pi i}{ka} \sum_{s=1}^{\infty} \frac{H_{\nu_s}^{(1)}(kr)\, H_{\nu_s}^{(1)}(kr')\, e^{i\nu_s|\varphi|}}{H_{\nu_s}^{(1)}(ka) \frac{\partial}{\partial \nu} \left[\frac{dH_{\nu}^{(1)}(ka)}{d(ka)} + igH_{\nu}^{(1)}(ka) \right]_{\nu=\nu_s}} \tag{1}$$

or a contour integral

$$\Gamma(r, \varphi', r') = \frac{\pi i}{2} \int_C e^{i\nu|\varphi|} H_{\nu}^{(1)}(kr') \left[H_{\nu}^{(2)}(kr) - \frac{\frac{dH_{\nu}^{(2)}(ka)}{d(ka)} + igH_{\nu}^{(2)}(ka)}{\frac{dH_{\nu}^{(1)}(ka)}{d(ka)} + igH_{\nu}^{(1)}(ka)} H_{\nu}^{(1)}(kr) \right] d\nu, \tag{2}$$

where the contour C encloses, in the positive direction, all the point ν_s (s = 1, 2, . . .) which are the roots of the equation

$$\frac{dH_{\nu}^{(1)}(ka)}{d(ka)} + igH_{\nu}^{(1)}(ka) = 0, \tag{3}$$

and which equation is determined from the impedance boundary condition

$$\frac{\partial\Gamma}{\partial r} + ikg\Gamma = 0 \text{ when } r = a \tag{4}$$

for the function Γ. In the interval (2) it is implied that $r' \geq r$, and when $r' < r$ we must interchange r and r'.

Formulas (1) and (2) yield a formal solution to our problem; we shall not stop to derive them, since the procedure is analogous to that used for the elliptical cylinder (see [2] and also [3] and [4]). Our purpose is an investigation of these formulas subject to the condition

$$ka \gg 1, \tag{5}$$

where asymptotic laws of diffraction on convex bodies enter into play.

In view of the fact that $r' \geq a$ and $r \geq a$, we can substitute for the Hankel functions in (1) and (2) their asymptotic expressions. The most general asymptotic expressions have the form

$$\begin{gathered} H_\nu^{(1)}(kr) = -i\sqrt{\frac{2}{\pi}}\sqrt[4]{\frac{-\tau}{(kr)^2-\nu^2}}\,w_1(\tau), \\ H_\nu^{(2)}(kr) = i\sqrt{\frac{2}{\pi}}\sqrt[4]{\frac{-\tau}{(kr)^2-\nu^2}}\,w_2(\tau), \end{gathered} \tag{6}$$

where $w_1(\tau) = w(\tau)$ and $w_2(\tau)$ are the Airy functions;

$$\tau = -\left(\frac{3\xi}{2}\right)^{\frac{2}{3}}; \qquad \xi = \frac{2}{3}(-\tau)^{\frac{3}{2}} \tag{7}$$

and

$$\begin{gathered} \xi = \int_\nu^{kr}\sqrt{1-\frac{\nu^2}{\rho^2}}\,d\rho = \sqrt{(kr)^2-\nu^2} - \nu \arcsin\sqrt{1-\frac{\nu^2}{(kr)^2}}, \\ \frac{\partial\xi}{\partial\nu} = -\arcsin\sqrt{1-\frac{\nu^2}{(kr)^2}}. \end{gathered} \tag{8}$$

These expressions can be found in the book by V. A. Fok [5]. We shall need in what follows the following particular cases of formulas (6), also found in [5]. If

$$\nu = ka + Mt, \quad M = \left(\frac{ka}{2}\right)^{\frac{1}{3}}, \quad |t| \ll M^2, \tag{9}$$

then formulas (6) assume the form

$$H_\nu^{(1)}(ka) = -\frac{i}{\sqrt{\pi}M}w_1(t), \quad H_\nu^{(2)}(ka) = \frac{i}{\sqrt{\pi}M}w_2(t). \tag{10}$$

On the other hand if

$$h = r - a \ll a, \quad y = \frac{kh}{M} = \frac{2M^2h}{a} \tag{11}$$

and condition (9) is satisfied, then

$$H_\nu^{(1)}(kr) = -\frac{i}{\sqrt{\pi}M}w_1(t-y), \quad H_\nu^{(2)}(kr) = \frac{i}{\sqrt{\pi}M}w_2(t-y). \tag{12}$$

For larger values of r (for example when $h \sim a$ or $h \gg a$), we can already use the Debye asymptotic formulas

$$H_\nu^{(1)}(kr) = \sqrt{\frac{2}{\pi}}\,\frac{e^{i\left(\xi-\frac{\pi}{4}\right)}}{\sqrt[4]{(kr)^2-\nu^2}}, \quad H_\nu^{(2)}(kr) = \sqrt{\frac{2}{\pi}}\,\frac{e^{-i\left(\xi-\frac{\pi}{4}\right)}}{\sqrt[4]{(kr)^2-\nu^2}}, \tag{13}$$

which are suitable when $\nu < kr$. On the other hand, if ν has the form (9), then (13) can be simplified in the following fashion: we put in the denominator simply $\nu = ka$, and calculate ξ from the formula

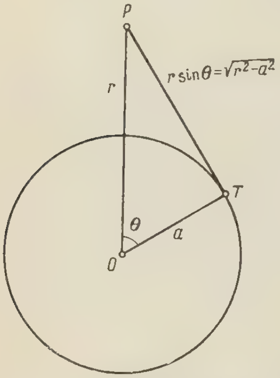


Fig. 1. Tangent PT to the circle of the cylinder, and the angle θ.

$$\xi(\nu) = \xi(ka) + \frac{\partial\xi}{\partial\nu}(ka)\,Mt. \tag{14}$$

As a result we obtain

$$H_\nu^{(1)}(kr) = \sqrt{\frac{2}{\pi kr\sin\theta}}\,e^{i\left(kr\sin\theta-\nu\theta-\frac{\pi}{4}\right)},$$

$$H_\nu^{(2)}(kr) = \sqrt{\frac{2}{\pi kr\sin\theta}}\,e^{-i\left(kr\sin\theta-\nu\theta-\frac{\pi}{4}\right)}, \tag{15}$$

where the angle θ is given by

$$a = r\cos\theta. \tag{16}$$

Its geometric meaning is clear from Fig. 1; this is the angle between the radius vector r = OP and the radius OT drawn to the "point of tangency" T, corresponding to the given point of observation P.

It is easy to show that the regions where (12) and (15) are valid overlap, since the formulas transform into each other continuously.

2. REFLECTION FORMULA

Let us assume that the contour C in the integral (2) can be drawn in such a way that the Debye formulas (13) can be applied to its principal portion. Then this integral assumes the form

$$\Gamma = i\int\limits_C \left[e^{i(\nu|\varphi|+\xi'-\xi)} + \frac{\sqrt{1-\left(\frac{\nu}{ka}\right)^2}-g}{\sqrt{1-\left(\frac{\nu}{ka}\right)^2}+g}\,e^{i(\nu|\varphi|+\xi'+\xi-2\xi_a)}\right] \times$$

$$\times \frac{d\nu}{\sqrt[4]{(kr')^2-\nu^2}\,\sqrt[4]{(kr)^2-\nu^2}}, \tag{17}$$

where ξ' and ξ_a are obtained from ξ with $r = r'$ and $r = a$, respectively. This integral can be calculated by the stationary-phase method (see Section 3 in [5]).

The first term in the square bracket of (17) has a phase

$$\omega(\nu) = \nu|\varphi| + \xi' - \xi, \tag{18}$$

the derivative of which

$$\omega'(\nu) = |\varphi| + \gamma' - \gamma, \tag{19}$$

where

$$\sin\gamma' = \frac{\nu}{kr'}; \quad \sin\gamma = \frac{\nu}{kr}. \tag{20}$$

The stationary-phase point ν_0 is obtained from the expression

$$\gamma - \gamma' = |\varphi| \quad \text{or} \quad \cos(\gamma - \gamma') = \cos\varphi \tag{21}$$

and is determined in explicit form by the formula

$$\nu_0 = \frac{kr'r}{\rho}\sin|\varphi|, \tag{22}$$

where

$$\rho = \sqrt{r'^2 - 2r'r\cos\varphi + r^2} \tag{23}$$

is the distance between the points P' and P. From Fig. 2 we can see the geometric meaning of the angles γ' and γ, corresponding to the value of ν_0, and it follows in particular that $\nu_0 = kH$.

We furthermore have

$$\begin{gathered} \omega(\nu_0) = \sqrt{(kr')^2 - \nu_0^2} - \sqrt{(kr)^2 - \nu_0^2} = k\rho, \\ \omega''(\nu_0) = -\frac{k\rho}{\sqrt{(kr')^2 - \nu^2}\sqrt{(kr)^2 - \nu^2}}, \end{gathered} \tag{24}$$

and the first term in (17) yields, after application of the stationary-phase method, simply the primary cylindrical wave

$$\Gamma^0 = \sqrt{\frac{2\pi}{k\rho}}\, e^{i\left(k\rho + \frac{\pi}{4}\right)}, \tag{25}$$

which is outgoing from the source P'.

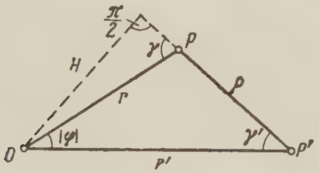


Fig. 2. Geometry of the direct ray.

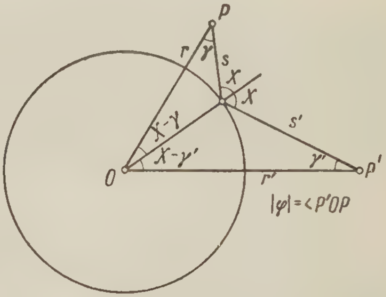


Fig. 3. Geometry of the reflected ray.

The second term in (17) has a phase

$$\omega(\nu) = \nu|\varphi| + \xi' + \xi - 2\xi_a, \tag{26}$$

the derivative of which

$$\omega'(\nu) = |\varphi| + \gamma' + \gamma - 2\chi, \tag{27}$$

where $\sin\chi = \nu/ka$, vanishes when

$$|\varphi| = 2\chi - \gamma' - \gamma. \tag{28}$$

The geometrical meaning of the angles γ', γ and χ, corresponding to the stationary-phase point ν_0, is illustrated in Fig. 3; in particular, χ is simply the angle of incidence, which equals the angle of reflection. It follows from Fig. 3 that

$$\omega(\nu_0) = \sqrt{(kr')^2 - \nu_0^2} + \sqrt{(kr)^2 - \nu_0^2} - 2\sqrt{(ka)^2 - \nu_0^2} = k(s' + s), \tag{29}$$

where $s' = P'N'$ is the length of the incident ray, $s = NP$ is the length of the reflected ray. In addition,

$$\omega''(\nu_0) = -\frac{1}{ka\cos\chi}\left(\frac{s'}{s'+a\cos\chi} + \frac{s}{s+a\cos\chi}\right). \tag{30}$$

The stationary phase method leads to the following expression for the reflected wave:

$$\Gamma^1 = \sqrt{\frac{2\pi}{kS}}\, e^{i\left[k(s'+s)+\frac{\pi}{4}\right]}\, \frac{\cos\chi - g}{\cos\chi + g}, \tag{31}$$

where

$$S = s' + s + \frac{2s's}{a\cos\chi}. \tag{32}$$

This expression is in full agreement with geometrical optics, and S can be rewritten

$$S = s'\left(1 + \frac{s}{s_0}\right), \tag{33}$$

where s_0 is the distance from the point of reflection N to the point of tangency of the continued ray NP with the imaginary caustic; s_0 is determined by the formula

$$\frac{1}{s_0} = \frac{2}{a\cos\chi} + \frac{1}{s'}. \tag{34}$$

If

$$g = \cos\chi_0, \tag{35}$$

then the expression (31) for the reflected wave vanishes when the angle of incidence is χ_0. However, the calculation with the aid of the more accurate method of stationary phase (which calls also for a more accurate formulation of Eq. (13)) shows that in this case transverse diffusion still leads to the formation of a reflected wave with an amplitude (for sufficiently large s' and s) has an order of magnitude which is ka times smaller than the amplitude of the wave reflected in other directions.

3. SURFACE WAVE

If the function $H_\nu^{(1)}(ka)$ is replaced by the asymptotic expression (10), and (12) is used, then Eq. (3) becomes

$$w_1'(t) - qw_1(t) = 0, \quad q = iMg. \tag{36}$$

Subject to the condition

$$g = -i|g|, \; q = M|g| \gg 1 \tag{37}$$

Eq. (36) has a "singular" root, which in the first approximation is

$$t = q^2 \tag{38}$$

and which in the second approximation has an exponentially small imaginary part. This singular root, as shown by Fok (Section 7 of [5]) is missing in the propagation of radio waves around the earth's surface, when

$$\frac{\pi}{4} < \operatorname{arc} q < \frac{\pi}{2}, \tag{39}$$

but if condition (37) is satisfied it does exist and corresponds to a surface wave propagating around the cylinder with low attenuation (see [7]). The dependence of this wave on the azimuth φ is determined in the first approximation by the factor $\exp\{i\nu|\varphi|\}$, where

$$\nu = ka\left(1 + \frac{1}{2}|g|^2\right). \tag{40}$$

Inasmuch as the z dependence of a surface wave propagating along an impedance plane (z is the coordinate along the plane) is determined by the factor

$$e^{ik\sqrt{1+|g|^2}\,z}, \tag{41}$$

we see from a comparison of (40) and (41) that Eqs. (36) and (38) can be used only when $|g| \ll 1$, when the phase velocity of the surface wave is close to the velocity of the waves in free space, and consequently the "surface character" of the wave is weak.

Elliott [6] gives certain numerical results for electromagnetic surface waves traveling over a corrugated cylinder or a cylinder with dielectric layer; these results have been obtained from a characteristic equation analogous to (3).

We note that the singular root (38) exists not only subject to condition (37), but also under the more general condition

$$g = -i|g|e^{i\psi}, \quad 0 \leqslant \psi \leqslant \frac{\pi}{6}, \quad |q| = M|g| \gg 1. \tag{42}$$

If in this case $|g| \gtrsim 1$, it is necessary to use in lieu of (36) a more complicated equation, obtained from the asymptotic formulas (6). In this case, and also under the condition (35), the parabolic equation in ray coordinates does not give a complete solution of our diffraction problem. If these cases are excluded, the exact solution (1) and (2) of the wave equation, as will be shown below, goes under condition (5) into the solution obtained in [1].

4. RIGOROUS SOLUTION IN RAY COORDINATES

Substituting into the contour integral (2) the asymptotic expressions (9), (10), and (12), and introducing the dimensionless variables

$$x = M|\varphi|, \quad y = \frac{kh}{M}, \quad y' = \frac{kh'}{M}, \tag{43}$$

we obtain for the function Γ the expression

$$\Gamma = \frac{2\pi i}{M} e^{ika|\varphi|}\Psi(x, y, y', q), \tag{44}$$

which is suitable under the condition that both points are near the cylinder ($h \ll a$, $h' \ll a$). The function Ψ is determined when $y' \geq y$ by the integral

$$\Psi(x, y, y', q) = \\ = \frac{1}{4\pi}\int_C e^{itx} w_1(t-y')\left[w_2(t-y) - \frac{w_2'(t) - qw_2(t)}{w_1'(t) - qw_1(t)} w_1(t-y)\right]dt, \tag{45}$$

taken over an infinite contour C, which encloses in the positive direction the poles t_s ($s = 1, 2, \ldots$) of the integrand, namely the roots of Eq. (36).

If the point P' is far from the cylinder, we must take for the function $H_\nu^{(1)}(kr')$ the asymptotic formula (15), and we obtain an expression

$$\Gamma = \sqrt{\frac{2\pi}{kr'\sin\theta'}}\, e^{i\left(kr'\sin\theta' + \frac{\pi}{4}\right)} e^{ika(|\varphi|-\theta)} V_1(z, y, q), \quad z = M(|\varphi| - \theta), \tag{46}$$

which agrees with Section 4 of reference [1].

In order to obtain the general formula (51), we use the fact that the regions where the asymptotic formulas (12) and (15) are valid overlap. We can therefore write them in the form of a single formula

$$H_\nu^{(1)}(kr) = -\frac{i}{\sqrt{\pi M}} e^{i\nu\left(\frac{r\sin\theta}{a} - \theta\right) - i\frac{2}{3}Y^{\frac{3}{2}}} w_1(t - Y), \tag{47}$$

where the variable

$$Y = M^2\left(\frac{r\sin\theta}{a}\right)^2 = M^2\frac{r^2-a^2}{a^2} \tag{48}$$

is equal to the variable y when $h \ll a$; in this case the angle θ is small, and the approximate relation

$$ka\left(\frac{r\sin\theta}{a}-\theta\right) = \frac{2}{3}Y^{\frac{3}{2}}, \tag{49}$$

is valid, and consequently the exponential factor in (47) is practically equal to unity, making this formula equivalent to the first formula of (12). If the variable Y is large, the function w_1 can be replaced by the asymptotic expression

$$w_1(t-Y) = \frac{1}{Y^{\frac{1}{4}}}e^{i\left(\frac{2}{3}Y^{\frac{3}{2}} - t\sqrt{Y} + \frac{\pi}{4}\right)}, \tag{50}$$

and formula (47) goes into the first formula of (15).

Substituting expressions of the form (47) into the series (1), we obtain a formula

$$\Gamma = \frac{2\pi i}{M}e^{i\left(k\sigma - \frac{2}{3}Y^{\frac{3}{2}} - \frac{2}{3}Y'^{\frac{3}{2}}\right)}\Psi(X, Y, Y', q), \tag{51}$$

which is the fundamental result of reference [1]. Here all the variables are determined by the ray coordinates derived in [11]; in polar coordinates, their values

$$\sigma = a(|\varphi| - \theta - \theta') + r\sin\theta + r'\sin\theta',$$
$$X = M\frac{\sigma}{a}, \quad Y' = M^2\left(\frac{r'\sin\theta'}{a}\right)^2. \tag{52}$$

The function Ψ is determined by the series

$$\Psi(x, y, y', q) = \sum_{s=1}^{\infty}\frac{e^{it_s x}}{t_s - q^2}\,\frac{w_1(t_s - y)}{w_1(t_s)}\,\frac{w_1(t_s - y')}{w_1(t_s)}, \tag{53}$$

which is equivalent to the integral (45).

In conclusion we note that even in the absence of a singular root (see Section 3), rather remote roots of Eq. (3) cannot be calculated by formulas (9) and (36). Therefore, in the illuminated region, where the series (53) converges poorly, formula (51) may result in a large error, and can be used only in the penumbral region — up to its joining the formulas of geometrical optics (25) and (31) for rays with sufficiently small slopes (see [1]).

REFERENCES

1. G.D. Malyuzhinets, L.A. Vaynshteyn. Radiotekhnika i elektronika, 1961, 6, 8, 1247.
2. L.A. Vainshtein, A.A. Fedorov. Radiotekhnika i elektronika, 1961, 6, 1, 31.
3. N.D. Kazarinoff, R.K. Ritt. IRE Trans., 1959, AP-7, December 21.
4. B.R. Levy, J.P. Keller. Canadian J. Phys., 1960, 38, 1, 128.
5. V.A. Fok. Diffraction of Radio Waves Around the Earth's Surface. Academy of Sciences Press, 1956.
6. R.S. Elliott, J. Appl. Phys., 1955, 26, 4, 368.
7. J.R. Wait, IRE Trans., 1960, AP-8, 4, 445.

Institute of Physics Problems,
Acoustics Institute,
Academy of Sciences, USSR

Received by editor 18 January 1961

INFLUENCE OF BENDING ON THE PROPERTIES OF PERIODICALLY DIAPHRAGMED WAVEGUIDES OF RECTANGULAR SECTION

B.N. Morozov

The effect of bending a comb-like slow-wave system on the properties of magnetic-type waves is considered. It is shown that the bending of a diaphragmed waveguide greatly influences the coupling resistance, and to a lesser extent, the phase velocity. Plots are given and a physical analysis is made of the results obtained.

In the investigation of comb-like slow-wave systems [2], particular attention is usually paid to a study of the properties of in-phase LE waves, since they have considerably greater coupling resistance compared with the out of phase waves. However, for uniformly bent slow-wave systems, which are used in vacuum tubes [3], and also in accelerators [4], this circumstance does not always hold true.

In a uniformly bent periodically diaphragmed waveguide (Fig. 1), the analogue of LE waves are the H_{mp} magnetic modes. In the case when the diaphragms are sufficiently closely spaced (i.e., the condition $\lambda \gg \theta R$ is satisfied), the field components of the H_{mp} waves in the interaction space have the form [5]

$$\begin{aligned}
E_r &= -A\frac{kn}{r}\sin\frac{\pi}{b}z\,[J_n(\varkappa r) - BN_n(\varkappa r)]\,e^{-jn\varphi},\\
E_\varphi &= -jAk\varkappa\sin\frac{\pi}{b}z\,[\dot{J}_n(\varkappa r) - B\dot{N}_n(\varkappa r)]\,e^{-jn\varphi},\\
E_z &= 0,\\
H_z &= A\varkappa^2\sin\frac{\pi}{b}z\,[J_n(\varkappa r) - BN_n(\varkappa r)\,e^{-jn\varphi},\\
H_r &= A\frac{\pi}{b}\varkappa\cos\frac{\pi}{b}z\,[\dot{J}_n(\varkappa r) - BN_n(\varkappa r)]\,e^{-jn\varphi},\\
H_\varphi &= jA\frac{n}{r}\frac{\pi}{b}\cos\frac{\pi}{b}z\,[\dot{J}_n(\varkappa r) - B\dot{N}_n(\varkappa r)]\,e^{-jn\varphi},
\end{aligned} \tag{1}$$

where $\varkappa = k\sqrt{1-(\lambda/2b)^2}$; $k = 2\pi/\lambda$; γ is the wavelength in free space; b is the height of the waveguide in the z direction; a is the half width of the interaction space; h_1 and h_2 are the heights of the diaphragm; $n = ckR/v_{ph}$ is the azimuthal propagation constant; R is the mean bending radius; c is the velocity of light; v_{ph} is the phase velocity of the wave on the radius R; m and p are the wave numbers of the field in the r and z directions respectively; J_n, $\dot{J}_n$, N_n, and $\dot{N}_n$, are the Bessel and Neumann functions of order n and their derivatives with respect to the argument; B is the right-hand (or the left-hand) part of the dispersion equation

$$\frac{\operatorname{tg}\varkappa h_1 J_n[\varkappa(R-a)] + \dot{J}_n[\varkappa(R-a)]}{\operatorname{tg}\varkappa h_1 N_n[\varkappa(R-a)] + \dot{N}_n[\varkappa(R-a)]} = \frac{\operatorname{tg}\varkappa h_2 J_n[\varkappa(R+a)] - J_n[\varkappa(R+a)]}{\operatorname{tg}\varkappa h_2 N_n[\varkappa(R+a)] - \dot{N}_n[\varkappa(R+a)]}. \tag{2}$$

In particular cases when the waveguide is loaded with diaphragms along a single wall, Eq. (2) simplifies. If the diaphragms are located along the internal bent wall ($h_2 = 0$), then Eq. (2) becomes

$$-\operatorname{tg}\varkappa h_1=\frac{\dot{J}_n[\varkappa(R+a)]\dot{N}_n[\varkappa(R-a)]-\dot{J}_n[\varkappa(R-a)]\,\dot{N}_n[\varkappa(R+a)]}{\dot{J}_n[\varkappa(R+a)]\,N_n[\varkappa(R-a)]-J_n[\varkappa(R-a)]\,\dot{N}_n[\varkappa(R+a)]},$$

and in (1) we have

$$B=\frac{\dot{J}_n[\varkappa(R+a)]}{\dot{N}_n[\varkappa(R+a)]};$$

if the diaphragms are located along the wall with the larger bending radius ($h_1 = 0$), then (2)

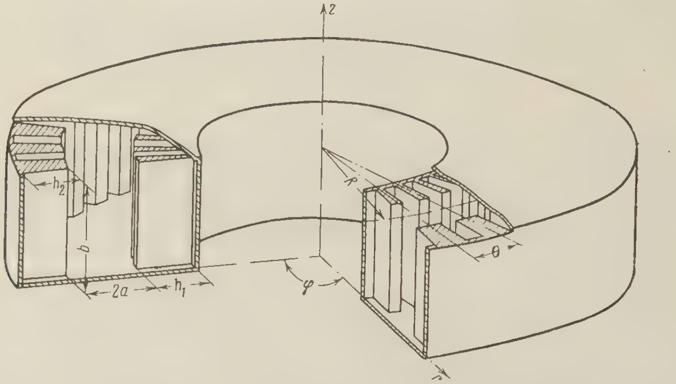


Fig. 1. Uniformly bent periodically diaphragmed waveguide.

assumes the form

$$\operatorname{tg}\varkappa h_2=\frac{\dot{J}_n[\varkappa(R+a)]\dot{N}_n[\varkappa(R-a)]-\dot{J}_n[\varkappa(R-a)]\dot{N}_n[\varkappa(R+a)]}{J_n[\varkappa(R+a)]\,\dot{N}_n[\varkappa(R-a)]-\dot{J}_n[\varkappa(R-a)]\,N_n[\varkappa(R+a)]},$$

where we have in (1)

$$B=\frac{\dot{J}_n[\varkappa(R-a)]}{\dot{N}_n[\varkappa(R-a)]}.$$

In an investigation of the effect of the bending on the properties of in-phase (H_{11}) and out of phase (H_{01}) waves in a waveguide loaded on both sides,* and type H_{01} waves in systems loaded on one side, it is advantageous to compare waveguides having identical width of the interaction space (ka = const) and identical slow-wave ratio c/v_{ph} on the central radius R. From the plots of Fig. 2, constructed in accordance with Eq. (2), it follows that the in-phase waves of a symmetrically loaded waveguide ($h_1 = h_2$) and the wave H_{01} of a waveguide with $H_2 = 0$ become accelerated, while the out-of-phase waves of the system with $h_1 = h_2$ and the lower type of the wave in the waveguide, loaded along the outer wall, are slowed down compared with the waves in the unbent waveguide of the same geometrical dimensions.

For waveguides loaded on one end, the effect of bending on the phase velocity of the wave is more considerable than for a symmetrically diaphragmed waveguide; curves 3 and 4 as shown by calculation, tend to a value determined from the following formula [6]

$$\frac{\lambda}{2b}\operatorname{th}\left(\frac{ka\,\lambda}{2b}\right)=\sqrt{1-(\lambda/2b)^2}\operatorname{tg}\varkappa h. \tag{3}$$

---

*The in-phase and out of phase waves of an unbent double "comb" [1, 2] are sometimes called in the literature symmetrical and antisymmetrical, respectively [3].

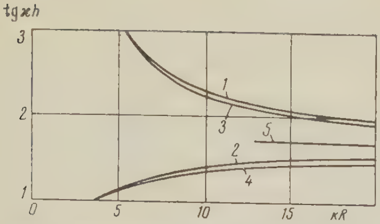


Fig. 2. Effect of bending on geometrical dimensions of waveguides, necessary to ensure constant slowing down ($ka = 2$, $\sqrt{1-(\lambda/2b)^2} = 0.5$, $c/v_{ph} = 1$):

1 — for in-phase wave of system with $h_1 = h_2$; 2 — for out of phase wave of system with $h_1 = h_2$; 3 — for H_{01} wave of system with $h_2 = 0$; 4 — for H_{01} wave of system of $h_1 = 0$; 5 — for straight comb, according to (3)

The fact that curves 1 and 3 or 2 and 4 are close to each other, can be readily explained from Fig. 3, which is based on formula (1). It turns out [7] that in a symmetrically loaded waveguide the in-phase waves "stick" to the internal surface, while the out-of-phase waves stick to the outer one. Therefore, in the former case the distribution of the azimuthal component of the electric field almost coincides with the distribution of the field in a waveguide loaded along the smaller-radius wall, while in the latter case the distribution of the azimuthal components of the electric field almost coincides with the distribution field in a waveguide loaded along the larger-radius wall. In the case of a symmetrically loaded waveguide, the resonators located along the wall with the large radius store very little of the in-phase wave energy, i.e., they have a very small influence on its slowing down; therefore, the curves 1 and 3 almost merge when the bending is strong. Consequently, from the point of view of efficiency of slowing down, out-of-phase waves of a symmetrically loaded waveguide and of a waveguide with diaphragms along the wall with the large radius are more suitable.

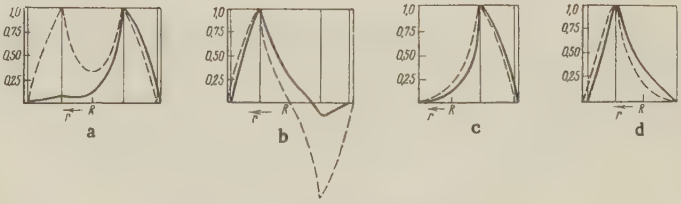


Fig. 3. Effect of bending on the distribution of E_φ in the transverse cross section for $kR = 8$;
a — in-phase wave of system with $h_1 = h_2$; b — out of phase wave of system with $h_1 = h_2$; c — h_{01} wave of system with $h_2 = 0$; d — H_{01} wave of system with $h_1 = 0$; dashed curves — for rectangular waveguide according to [2]

The coupling resistance is determined from the relation

$$R_c = \frac{E_\varphi^2}{2\beta^2 P}, \tag{4}$$

where $\beta = 2\pi/\lambda_w$; P is the energy flux through the waveguide interaction space, averaged over the cycle. Substituting in (4) the expressions for the field components from (1) we obtain (in the practical system of units) for the average radius R, on which $c/v_{ph} = 1$,

$$R_c = \frac{480\pi}{kb} \frac{\dot{Z}_n^2(R)}{f_n(R-a) - f_n(R+a)}, \tag{5}$$

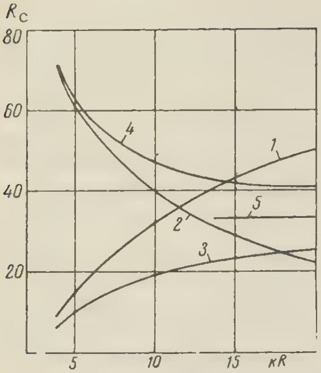
where

$$\dot{Z}_n(R) = \dot{J}_n(\varkappa R) - B\dot{N}_n(\varkappa R)$$

and according to [8],

$$f_n(R \pm a) = Z_0^2\,[\varkappa(R \pm a)] + 2\sum_{s=1}^{n-1} Z_s^2\,[\varkappa(R \pm a)] + Z_n^2\,[\varkappa(R \pm a)]$$

Formula (5) was used to calculate the curves of Fig. 4. Curve 1 tends with increasing R to the value $R_c = 90$ ohms, which is determined from the relation for the unbent waveguide



$$R_c = 480\left(\frac{\lambda}{2b}\right)^2 \frac{\pi^2\varkappa^2}{b^2\left(\operatorname{sh} 2\frac{\pi a}{b} - 2\frac{\pi a}{b}\right)}\,.$$

Fig. 4. Effect of bending on the coupling resistance ($ka = 2$, $\sqrt{1-(\lambda/2b)^2} = 0.5$, $c/v_{ph} = 1$:

1 — for in-phase wave of system with $h_1 = h_2$; 2 — for out of phase wave of system with $h_1 = h_2$; 3 — for H_{01} wave of system with $h_2 = 0$; 4 — for h_{01} wave of system with $h_1 = 0$; 5 — for straight comb according to (6).

As the waveguide is bent, the coupling resistance of the out-of-phase waves of a symmetrically loaded waveguide increases to a certain value R (in the case of strong bending) exceeding the coupling resistance of the in-phase waves, which decreases with increased bending. An analogous situation obtains in the case of asymmetrically loaded waveguides. In the latter case, however, when the waveguide is straightened out, the coupling resistance does not drop to zero as it does for curve 2, but tends to the value

$$R_c = 480\left(\frac{\lambda}{2b}\right)^2 \frac{\pi^2\varkappa^2 \operatorname{sh}^2 \pi\frac{a}{b}}{b^2\left(\operatorname{sh} 2\frac{\pi a}{b}\operatorname{ch} 2\frac{\pi a}{b} + 2\frac{\pi a}{b}\right)} \tag{6}$$

for a rectangular comb. From the physical point of view, this phenomenon is explained by the redistribution of the field in the transverse cross section when the waveguide is bent (Fig. 3).

A comparison of Figs. 2 and 4 shows that the bending greatly influences the coupling resistance and influences to a lesser extent the geometrical dimensions of the waveguide (for constant slowing down on the mean radius). Thus, for example, in the case of in-phase waves of a symmetrically loaded waveguide with $kR = 10$, the coupling resistance differs by 60% from the unbent waveguide, while h differs by 5%.

Consequently, in the case of strong bending, the out-of-phase waves of a symmetrically loaded waveguide and a type H_{01} wave of a waveguide with diaphragms along the larger-radius wall can be of definite interest. We note, in addition, that the out of phase waves have a smaller dispersion [1], and the critical wavelength of the H_{01} mode is greater than that of the other types of waves, which greatly facilitates the struggle against parasitic oscillations.

We can conclude from the foregoing that when waveguides with large curvature are used ($kR \simeq 15$), it is advantageous to use the out-of-phase waves of a symmetrically loaded waveguide (when $kR < 10$, in the particular case under consideration), and the H_{01} waves of a waveguide with diaphragms along the larger-radius wall ($kR < 15$). In the case of weak bending, however, ($kR > 20$), the most effective coupling of the electron beam can be attained only with the in-phase wave of a symmetrically loaded waveguide.

The author is grateful to L. V. Chikulina for help in the calculations.

REFERENCES

1. N.V. Trunova. On the theory of in-phase waves in a rectangular waveguide with periodic comb-like structure. Izv. Vuzov MVO SSSR (Radiotekhnika), 1958, 1, 1, 105.
2. E.S. Kovalenko, V.I. Shimanskii. In-phase waves in a diaphragmed waveguide of rectangular cross section. Ibid 1960, 3, 2, 153.
3. E.S. Grishin, O.N. Aristarkhova, I.Sh. Beluga, A.K. Bryukhova, I.Ya. Pinchuk, R.A. Sinin, Kh. I. Spektor, E.P. Teveleva, I.N. Khlebnikov. Amplifying and generating traveling wave tubes of the magnetron type. Transactions of the Scientific Research Institute for the Ministry of Radio Industry, USSR, 1957, 12, (48), 118.
4. A.A. Vorob'ev, A.N. Didenko, E.S. Kovalenko. Acceleration of electrons in a traveling-wave cyclic waveguide accelerator. Atomnaya Energiya, 1960, 8, 4, 459.
5. B.N. Morozov. Certain problems in the propagation of electromagnetic waves in bent

waveguides loaded with periodic structures. Izv. Vuzov MVO SSSR, Radiotekhnika, 1960, 3, 4, 493.
6. R.M. Leont'eva. Longitudinal electric waves in a periodic diaphragm rectangle waveguide. Radiotekhnika, 1957, 12, 12, 36.
7. A.N. Didenko. Propagation of electromagnetic waves in bent loaded waveguides. Radiotekhnika i elektronika 1959, 4, 2, 172.
8. G.N. Watson. A Treatise on the Theory of Bessel functions. (Russian translation), IL, 1949, Part 1, p. 150 (Cambridge University Press, 1944).

Received by editor 21 November 1960

INTERACTION BETWEEN AN ELECTRON BEAM AND THE FIELD OF A HELIX IN THE REGION OF AN ABSORBING INSERT

Yu. N. Pchel'nikov

The propagation constant of a growing wave in a helix with absorption, in the presence of an electron beam is calculated approximately. The power dissipated in the region of the absorbing insert of a traveling wave tube is calculated, and recommendations are made on the choice of parameters and position of the insert.

INTRODUCTION

Theoretical and experimental investigations of the phenomena occurring in traveling wave tubes (TWT) in the vicinity of an absorbing insert have been the subject of many papers [1-6] the results of which indicate that the position and parameters of the absorbing insert exert an appreciable influence on the operation of the TWT. However, the foregoing investigations do not display fully the physics of the phenomena that take place in the region of the absorbing insert, nor do they propose a procedure for correct determination and choice of the parameters and position of the absorber. In reference [7] is considered the distribution of the high-frequency field along a model of a TWT in the low-signal mode, as a function of the position attenuation parameter of the absorbing insert. On the other hand, an estimate of the influence of the insert on the maximum output power of the TWT has not yet been made.

We give below an approximate calculation of the field of an increasing wave in the region of the insert, an estimate of the dissipated power, and an evaluation of the dependence of the efficiency of a tube with helical slow-wave system on the attenuation inserted and on the location of the absorbing insert.

Determination of the amplitude of the electromagnetic field in the region of the absorbing insert calls for calculation of the propagation constants of electromagnetic waves in a helix with sufficiently large absorption, in the presence of an electron beam. In the general case, the solution of such a problem entails great difficulties, owing to the need of solving transcendental equations that contain Bessel functions of complex arguments. On the other hand, the use of various models which deviate greatly from real systems does not give correct results.

The problem can be greatly simplified by confining oneself to an examination of a helical line with such parameters, that the Bessel functions contained in the dispersion equation of the system have sufficiently large arguments (in practice, greater than two) and can be replaced by simpler approximations based on asymptotic expansions. If we consider at the same time the case when the beam fills all of the internal part of the helix i.e., if we set the radius of the beam equal to the radius of the helix, then the dispersion equation becomes algebraic.

A similar or nearly equivalent case can take place also in real systems, and the results obtained by analyzing such a case make it possible to estimate, if not quantitatively at least qualitatively, the influence of the absorbing insert of a TWT with a different beam and helix geometry.

It is known that attenuation in a helical line can be due both to the presence of absorbing material in the medium surrounding the helix, and to the direct shorting of the helix turns by a layer having a certain resistance. To simplify the calculation, the first method of producing an absorbing insert will be taken into account by using a model in which the helix is surrounded on the outside by an unbounded absorbing medium, while the second method will comprise the analysis of a helix having a certain conductivity in a direction perpendicular to the direction of the helix turns.

1. DERIVATION OF SIMPLIFIED DISPERSION EQUATION

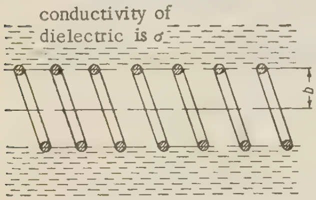


Fig. 1. Helix surrounded by unbounded absorbing medium

Let us consider (Fig. 1) a helix of radius b with an electron beam of the same radius, surrounded on the outside by an unbounded absorbing medium, having a certain conductivity σ. The dielectric constant of such an absorbing medium can be represented by the complex quantity

$$\varepsilon_1 = \varepsilon\left(1 - j\frac{\sigma}{\varepsilon\omega}\right), \tag{1}$$

where ϵ is the dielectric constant of the outer medium without account of conductivity, ω is the angular frequency of the field and $j = \sqrt{-1}$.

Let the dielectric constant of the medium inside the helix be denoted by ϵ_0, and let the permeability of the medium be μ_0 in both regions.

We seek the values of the propagation constants of the electromagnetic waves in this system, confining ourselves to axially-symmetrical solutions. We assume henceforth that the electromagnetic field intensity components depend on the time and on the longitudinal coordinate z as $\exp(j\omega t + \gamma z)$, where γ is the propagation constant, defined by

$$\gamma = -j\beta + \alpha. \tag{2}$$

Here β is the phase constant and α is the wave increment constant. Designating the values of the constants in the absence of a beam by the index 0, we obtain

$$\gamma_0 = -j\beta_0 - \alpha_0, \tag{3}$$

where α_0 is the attenuation constant.

Solving the problem subject to the customary idealization (replacing the helix by an anisotropically conducting cylinder), we obtain a transcendental dispersion equation for the propagation constants in the system under consideration. The resultant equation contains modified Bessel functions of the first and second kind, of the arguments bT, $b\tau$, and $b\tau_1$, where T, τ, and τ_1 are radial constants, of which the first two are in the internal region and the third outside the helix; τ and τ_1 are connected with the electromagnetic wave propagation constant γ by the relations

$$\gamma^2 = -\tau^2 - k^2 = -\tau_1^2 - k_1^2, \tag{4}$$

where k and k_1 are the phase constants, determined by the usual relations: $k = \omega\sqrt{\epsilon_0\mu_0}$, $k_1 = \sqrt{\epsilon_1\mu_0}$.

Since in most cases the slowing down in helical lines is sufficiently large, we can neglect k^2 and k_1^2 compared with τ^2 and τ_1^2, and it follows from (2) that

$$-j\tau_1 \simeq -j\tau \simeq \gamma. \tag{5}$$

The constants T and τ are related by

$$\frac{T^2}{\tau^2} = 1 + \frac{q}{\left(\gamma + j\,\frac{\omega}{u_0}\right)^2}, \tag{6}$$

where q is the square of the plasma wave number of an infinitely broad beam of the same density as the considered beam; u_0 is the constant component of the electron velocity.

As stated earlier, at sufficiently large values of the arguments, the modified Bessel functions and their ratios, contained in the dispersion equation, can be replaced by their asymptotic expansions. For this purpose we can use the following approximate relations:

$$\begin{gathered}\frac{I_0(X)}{I_1(X)} \simeq 1 + \frac{1}{2X}, \quad \frac{I_1(X)}{I_0(X)} \simeq 1 - \frac{1}{2X}, \\ \frac{K_0(X)}{K_1(X)} \simeq 1 - \frac{1}{2X}, \quad \frac{K_1(X)}{K_0(X)} \simeq 1 + \frac{1}{2X}, \\ I_0(X)\,K_0(X) \simeq I_0(X)\,K_1(X) \simeq I_1(X)\,K_0(X) \simeq I_1(X)\,K_1(X) \simeq \frac{1}{2X}.\end{gathered} \tag{7}$$

Here $I_0(X)$, $I_1(X)$, $K_0(X)$, $K_1(X)$ are the modified Bessel functions of the corresponding argument.

After certain transformations with account taken of (5) and (7), the resultant transcendental dispersion equation of our system is transformed to the following algebraic equation

$$\frac{2\gamma^2}{k^2\,\mathrm{tg}^2\,\Phi} + \frac{\varepsilon_1}{\varepsilon_0} + 1 = 1 - \frac{T}{\tau}, \tag{8}$$

where Φ is the angle between the direction of the turns and the longitudinal axis of the helix.

In the absence of an electron beam, q vanishes, and as follows from (6) $T = \tau = \tau_0$, where τ_0 is the radial constant in a system without beam.

In order not to complicate the problem, we shall assume henceforth the real part of the dielectric constant ϵ_1 to be ϵ_0. Then

$$\frac{\varepsilon_1}{\varepsilon_0} = 1 - j\,\frac{\sigma}{\omega\varepsilon_0}. \tag{9}$$

Putting $T = \tau$ in (8) we obtain, with account taken of (9), the following expressions for γ_0, β_0, and α_0:

$$\gamma_0 = -k\,\mathrm{tg}\,\Phi\sqrt{1 - \frac{j\sigma}{2\varepsilon_0\omega}}, \tag{10}$$

$$\beta_0 = \sqrt{k^2\,\mathrm{tg}^2\,\Phi + \alpha_0^2}, \tag{11}$$

$$\alpha_0 = \frac{\sigma}{4\beta_0\varepsilon_0\omega}. \tag{12}$$

It follows from (11) that with increasing attenuation constant α_0, the phase constant β_0 increases, i.e., the phase velocity of the wave decreases.

2. DERIVATION OF THE CHARACTERISTIC EQUATION

By eliminating T and τ from (8) and (6) and using (10) we obtain an equation for the propagation constant γ:

$$(\gamma^2 - \gamma_0^2)\left(\gamma + j\,\frac{\omega}{u_0}\right)^2(\gamma^2 - \gamma_0^2 - k^2\,\mathrm{tg}^2\Phi) = \frac{q}{4}\,k^4\,\mathrm{tg}^4\,\Phi. \tag{13}$$

The resultant equation has six roots, two of these (in the third factor) are close to zero and have no physical meaning, so that they can be excluded. We know that in the amplification mode of the TWT the propagation constants of interest to us are nearly equal to β. Using this fact, and also the smallness of q, we derive by approximate successive transformations the following fourth-degree equation for γ:

$$(\gamma^2-\gamma_0^2)[(\gamma+j\beta_e)^2+\Gamma q] = -\frac{q}{4}(\beta_0^2-\alpha_0^2), \tag{14}$$

where

$$\Gamma = \frac{1}{4}\,\frac{1-\frac{\alpha_0^2}{\beta_0^2}}{\eta^2+\frac{2j\alpha_0}{\beta_0}}; \qquad \eta = \frac{\omega}{u_0\beta_0}. \tag{15}$$

This equation is the usual characteristic TWT equation and finds the propagation constant of four waves: three forward waves and one backward wave. The values of these waves are roughly close to the values of the roots

$$\gamma_{1,2}\simeq\pm\gamma_0, \quad \gamma_{3,4}\simeq -j\left(\frac{\omega}{u_0}\pm\sqrt{\Gamma q}\right). \tag{16}$$

The first two roots determine the propagation constant of the forward and backward waves in the system without beam, while the third and fourth are the propagation constants of the slow and fast waves of the beam.

A solution of (14) shows that in the range where η is close to unity, there exists in a helical line surrounded by an absorbing medium a growing wave whose velocity is close to the velocity of the slow wave of the beam. The region of growing solutions is unbounded on the $\eta > 1$ side.

An ordinary absorbing insert has an attenuation corresponding to $\alpha_0/\beta_0 \simeq 0.2$. At this value of $\alpha_0\beta_0$, the roots of Eq. (4) separate, and analytic solutions can be found for them. Calculations have shown that the results differ little from those defined by (16).

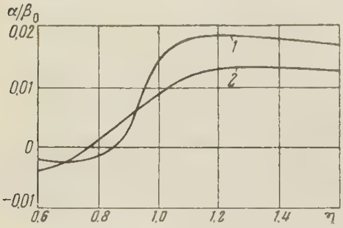


Fig. 2. Dependence of the relative value of the build-up constant α/β_0 on the parameter η for two values of the attenuation parameter: 1 – $\alpha_0/\beta_0 = 0.2$; 2 – $\alpha_0/\beta_0 = 0.4$.

Fig. 2 shows the dependence of the build-up constant of the slow wave of the beam on the parameter η for two values of α_0/β_0, namely 0.2 and 0.4. The value of q/β_0^2 was set at 10^{-2}, corresponding to relatively large beam-current densities. It can be readily verified that the maximum build-up constant in a system with losses corresponding to $\alpha_0/\beta_0 = 0.2$ is found here to be one order of magnitude less than the build-up constant in a lossless helical line.

3. HELIX WITH FINITE SURFACE CONDUCTIVITY

A calculation similar to that given above was made for a helical line with electron beam (Fig. 3), differing in that the absorption in the line was due to surface conductivity σ_1 in a direction perpendicular to that of the helix terms. The resultant dispersion equation has the form

$$\frac{k^2\,\mathrm{tg}^2\,\Phi}{\gamma^2}+1+\frac{\omega\mu_0\sigma_1}{\gamma}(\mathrm{tg}^2\,\Phi-1) = \frac{\frac{T}{\tau}-1}{\frac{T}{\tau}+1}\left[1+\frac{\omega\mu_0\sigma_1}{\gamma}(\mathrm{tg}^2\,\Phi-1)\right]. \tag{17}$$

In the absence of a beam $T/\tau = 1$ and (17) is transformed into the "cold" dispersion equation of a system relative to the propagation constant γ_0. In this case β_0 and α_0 are defined by

$$\beta_0 = \sqrt{k^2\,\mathrm{tg}^2\,\Phi-\alpha_0^2}, \tag{18}$$

$$\alpha_0 = \frac{\mathrm{tg}^2\Phi-1}{2}\,\omega\mu_0\,\sigma_1. \tag{19}$$

Unlike the phase constant in a helix surrounded by an unbounded absorbing medium, the phase constant in this system decreases and consequently the velocity of the wave increases with increasing attenuation constant α_0, up to a value $\alpha_0 = k \tan \Phi$. This is due to the increase in the coupling between the neighboring turns with increasing conductivity σ_1. However, at real values of the attenuation, the velocity of the wave hardly differs in either case from the value it has in the absence of attenuation.

The foregoing calculation enables us to draw the important conclusion that in the region of the absorbing insert the amplitude, and consequently the power of the slow wave of the beam, in spite of the presence of absorption, not only do not decrease, but even increase. In this case the radial structure of the field of the amplified wave hardly differs from the structure of the field of the wave in the insert region in the absence of a beam. This can be readily verified by comparing the real parts of the radial constants T and τ with the real part of the constant τ_0. It can be shown that the power fluxes in this system are determined by the real parts of the radial constants, and consequently the ratio of the power flux passing along the absorbing insert to the power flux scattered in the absorber is the same for the slow wave of the electron beam as for the natural wave of the system, i.e., it is determined by the value of α_0.

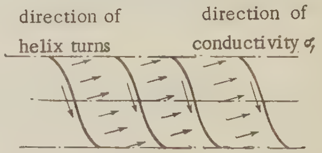


Fig. 3. Helix with finite conductivity in a direction perpendicular to the direction of the turns.

The fact that in spite of considerable absorption the power of the beam wave does not decrease, and even increases somewhat, is due to the conversion of the energy of the electron beam into high frequency energy of the electromagnetic field.

Because of the smallness of the build-up constant and the insignificant length of the insert, the power of the beam can be assumed the same at the beginning and the end of the absorbing insert. Since it is precisely the amplitude of the slow wave of the beam that increases in the active portions of the tube, then if the ends of the insert are so matched that no transformation of waves takes place, it can be assumed that the absorbing insert causes no attenuation of the amplified signal.

4. DETERMINATION OF DEPENDENCE OF THE ELECTRONIC EFFICIENCY ON THE POSITION OF THE INSERT AND ON THE ATTENUATION IT PRODUCES

The results obtained enable us to determine approximately the energy scattered in the absorbing insert and the dependence of the electronic efficiency of the tube on the position of the insert and on the attenuation it introduces. The power loss in an element dz of the absorbing insert is determined by the expression

$$dP_a = 2\alpha_0 P(z)\,dz, \tag{20}$$

where P(z) is the wave power in the element under consideration. Assuming the P(z) to be constant along the insert and denoting it by P_1, we find that the power P_a scattered in the insert is given by

$$P_a = 2\alpha_0 P_1 l, \tag{21}$$

where l is the length of the insert. If we denote by L the total attenuation (in decibel) introduced by the insert, we get

$$P_a = P_1 \frac{L}{4.34}\,. \tag{22}$$

Since usually L is on the order of 40-50 db, it follows from (22) that the power loss can exceed by a factor of 10 the power of the transmitted wave.

It is known that the maximum output power of the tube, or the saturation power (P_{out}^S) is determined by the power of the electron beam (P_0) and by the electronic efficiency (η_e). The latter is determined by the parameters of the slow-wave system and of the beam, including the losses in the slow-wave system and in the insert:

$$\eta_e = \frac{P_{out}^s}{P_0}. \qquad (23)$$

If we neglect the losses in the slow-wave system, then for given system and beam parameters, η_e will decrease with increasing power absorbed by the insert. Assuming the maximum power given up by the beam is independent of which part is scattered in the absorber and which goes into the useful signal, we can write

$$P_{out}^s = P_0 \eta_{max} - P_a, \qquad (24)$$

where η_{max} is the maximum value of the electronic efficiency.

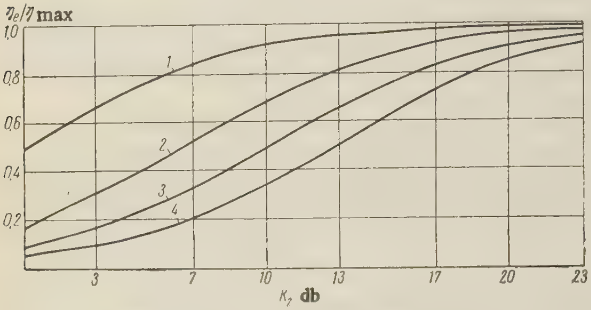


Fig. 4. Dependence of relative value of electronic efficiency, η_e/η_{max}, on the gain K_2 of the output portion by different values of insert absorption L: 1 — L = 4.34; 2 — L = 20; 3 — L = 43.4; 4 — L = 80

As follows from (22), the power P_a is proportional to the power of the wave in the region of the insert, P_1. If we denote by K_2 the gain of the output portion of the tube, then P_1 is given by

$$P_1 = \frac{P_{out}^s}{K_2}. \qquad (25)$$

Using (24) and (25), we find that the ratio of the saturation power to the maximum output power or, what is the same, the ratio η_e/η_{max} is

$$\frac{\eta_e}{\eta_{max}} = \frac{1}{1 + \frac{L}{4.34 K_2}} \qquad (26)$$

Fig. 4 shows plots of η_e/η_{max} vs. K_2, calculated from formula (26) for different values of L. The gain K_2 is indicated in decibels. Since its value is approximately proportional to the length of the output portion of the tube, the curves given illustrate the dependence of the saturation power on the position of the insert and on the attenuation L introduced by it. The greater K_2 and the less L, the greater η_2.

The maximum value of K_2 is limited by the excitation of the tube at the output portion, and consequently is determined by the quality of matching of its ends (the absorber and the output matching device). The minimum value of L is limited by the excitation of the entire tube and should exceed by 10 db the measured value of the gain. Precisely these considerations should govern the position and parameters of the absorbing insert. The length of the initial portion of the tube is determined by the over-all gain and must not exceed the length of the output portion.

In conclusion we note that the calculation offered here provides only a qualitative estimate of the phenomena occurring in the region of the absorbing insert. The results obtained cannot be used when the phase velocity of the electromagnetic waves in the region of the absorbing insert differs greatly from the velocity in the remaining portions of the tube.

The author is grateful to L.N. Loshakov for a discussion of the present paper.

REFERENCES

1. C.C. Cutler, D.J. Brangaccio. IRE Trans., 1953, PGED-3, 9.
2. J.J. Caldwell, IRE Trans., 1953, PGED-4, 28
3. J.J. Caldwell, O.Z. Hoch. IRE Trans., 1956, ED-3, 1, 6.
4. J.E. Rowe, Proc. IRE, 1956, 44, 2, 200.
5. C.H. Dix. IRE Trans., 1960, MTT-8, 1, 121.
6. C.K. Birdsall et al. IRE Trans., 1959, ED-6, 1, 6.
7. A.M. Kats. Radiotekhnika i Elektronika, 1960, 5, 12, 1986.

Received by editors 28 December 1960

EXCITATION OF A DELAY LINE BY MEANS OF A PERIODIC SEQUENCE OF IDEALLY FOCUSED ELECTRON BUNCHES

G. F. Filimonov

An analytic solution is found for the problem of excitation of a homogeneous (and also inhomogeneous if $\epsilon \to 0$) delay line by means of a periodic sequence of ideally focused (i.e., with zero transit angle) electron bunches. Numerical values are obtained for the electronic efficiency of the tube and for its optimal length. The results are compared with calculations made for different methods of exciting the delay line.

INTRODUCTION

Many recent papers (see, for example, [1]) are devoted to theoretical and experimental investigations of traveling wave tubes. These treat extensively the problem of the interaction of the slow-wave field with an electron beam which has at the entrance to the delay line a constant velocity and a constant density; the maximum attainable efficiencies are obtained as functions of several TWT parameters, primarily the smallness parameter ϵ and the velocity detuning parameter ξ_0.

Certain interest has also been displayed recently in investigations of properties of tubes in which the UHF waves are generated by a periodic sequence of focused electron bunches. The present work is aimed at a theoretical investigation of the energy capabilities of a tube of this kind, constructed in the form of a delay line the input to which is a nonrelativistic electron beam with periodically varying (in accordance with a certain law) velocity and density.

The field excited will depend, naturally, both on the duration (i.e., the "transit angle") of the electron bunches entering the tube, and the dispersion of the electron velocities in the bunches. It is possible to single out here, however, a simpler problem, which is more amenable to analytic solution, that of excitation of a delay line by a periodic sequence of ideally focused electron bunches (i.e., bunches with zero transit angles). This article is devoted to a solution of this problem.

1. FORMULATION OF THE PROBLEM

The fundamental equations of the nonlinear TWT theory can be written in the form given, for example, in [2]:

$$-\frac{\partial^2 t}{\partial z^2} = \left(\frac{\partial t}{\partial z}\right)^3 \operatorname{Re} \frac{e}{m} E e^{-i\omega t}, \tag{1}$$

$$\frac{d^2 E}{dz^2} + h_0^2 E = -i h_0 R_0 j, \tag{2}$$

$$j = 2 \int_0^T \frac{dt_0}{T} j_0(t_0)\, e^{i\omega t_0}, \tag{3}$$

where t_0 is the time when the electron enters the tube; z, $t = t(z, t_0)$ are its running coordinates and time; e and m are the charge and mass of the electrons; E is the complex amplitude of the field acting on the electron (the Coulomb field is discarded by virtue of the assumption that the bunches are ideally focused); h_0 is the wave number of the slow wave; R_0 is the coupling coefficient; T is the period and $j_0(t_0)$ is the electron current at the input to the tube. In the ordinary TWT theory this current is assumed constant ($j_0(t_0) = j_0 = \text{const}$);

in the case which we are considering, that of ideal focusing, $j_0(t_0)$ is proportional to the δ- function:

$$j_0(t_0) = j_0 \delta\left(\frac{t_0}{T}\right), \tag{4}$$

where $j_0 = \int\limits_0^T \frac{dt_0}{T} j_0(t_0)$ is the electron current averaged over the cycle.

In Eqs. (1) — (3) it is convenient to change over to dimensionless quantities, by putting

$$\omega t = \omega t_0 + h_0 z + w(\zeta), \quad z = \frac{v_0 \zeta}{\varepsilon \omega}, \quad h_0 = h_{00} = \frac{\omega}{v_0}(1 + \varepsilon \xi_0),$$
$$\varepsilon^3 = \frac{R_0 j_0}{4\left(\frac{\omega}{v_0}\right)^2 V_0}, \quad E = \frac{m}{e} v_0 \omega \varepsilon^2 f. \tag{5}$$

Here v_0 is the velocity of the electron at the entrance to the delay line; V_0 is their entrance potential; ϵ is the smallness parameter. Assuming further

$$f = F_+(\zeta) e^{ih_0 z} + F_-(\zeta) e^{-ih_0 z} \tag{6}$$

with the additional condition

$$F'_+(\zeta) e^{ih_0 z} + F'_-(\zeta) e^{-ih_0 z} = 0, \tag{7}$$

we obtain in lieu of (1) — (3) the following system of dimensionless equations, which describe the nonlinear processes in a TWT excited by a periodic sequence of ideally focused electron bunches:

$$-w'' = (1 + \varepsilon \xi_0 + \varepsilon w')^3 \operatorname{Re} (F_+ e^{ih_0 z} + F_- e^{-ih_0 z}) e^{-i\omega t}, \tag{8}$$
$$F'_\pm = \mp 2 e^{i(h_0 z \mp h_0 z + w)}. \tag{9}$$

The prime denotes here differentiation with respect to ζ.

According to the derivation given in [3] for the field Eq. (2), this equation determines the amplitude of only the "resonant" natural mode of the delay line, the phase velocity of which is close to v_0. All the remaining natural modes of the delay line either pertain to the field of the repulsion forces (slow waves) or are left out (fast waves) since it is assumed that they produce small oscillations, which change insignificantly the principal solution defined by (1) — (3). Therefore, integrating approximately Eq. (9) for F_-, we obtain

$$\operatorname{Re} F_- e^{-ih_0 z - i\omega t} \simeq \operatorname{Re} d e^{-2ih_0 z - iw} - \frac{\varepsilon^3 w''}{8}. \tag{10}$$

Here d is the integration constant. It is easy to see [see (8)] that the "backward" wave F_- causes, in the case of excitation of the delay line by a periodic sequence of ideal bunches, only small oscillations of the solution (accurate to effects $\sim \epsilon^3$), and within the limits of the accuracy of the method this wave should be left out in the equation of motion (8). When $F_1 = 0$, we obtain for the power flux of the high-frequency field the following expression

$$P = \eta j_0 V_0; \quad \eta = \varepsilon p; \quad p = \frac{|F_+|^2}{2}. \tag{11}$$

Here η is the efficiency of the tube and p its dimensionless power.

2. SOLUTIONS OF EQS. (8) AND (9)

After all the simplifications are made, we obtain for solution the following system of equations:

$$-w'' = (1 + \varepsilon \xi_0 + \varepsilon w')^3 \operatorname{Re} F_+ e^{-iw}, \tag{8a}$$
$$F'_+ = -2 e^{iw}. \tag{9a}$$

Introducing the amplitude and phase of the field

$$F_+(\zeta) = A(\zeta)\, e^{i\varphi(\zeta)} \tag{12}$$

and calculating the first integrals of (8a) and (9a), we obtain the following equations

$$\varepsilon w'_\zeta = (1-\eta)^{-1/2} - 1 - \varepsilon\xi_0, \tag{13}$$

$$\varphi' A = \pm\sqrt{4 - A'^2}, \tag{14}$$

$$\varepsilon\varphi' = (1+\sqrt{1-\eta})^{-1} - \frac{1+\varepsilon\xi_0}{2}, \tag{15}$$

which actually solve our problem. Indeed, substituting (14) in (15) and putting $A' = 0$, we obtain the algebraic equation

$$\frac{\varphi'}{|\varphi'|}\sqrt{\frac{2\varepsilon^3}{\eta}} = \frac{1}{1+\sqrt{1-\eta}} - \frac{1+\varepsilon\xi_0}{2}, \tag{16}$$

from which we determine the maximum efficiency of the instrument (for given ϵ and ξ_0), and after solving (14) and (15) with respect to A' and integrating we obtain the point $\zeta = \zeta_m$, at which this maximum is reached. In the derivation of (13) — (15) we used as boundary conditions the equations

$$w(0) = 0;\quad w'(0) = -\xi_0;\quad A(0) = 0, \tag{17}$$

which correspond to the initial velocity of the electron equal to v_0, and to absence of the field of the "forward" wave at the entrance to the tube. Eq. (16) is best solved by making the substitution

$$\eta = \sin^2\psi,$$

and for the sake of being definite we assume ψ to lie in the first quadrant ($0 \leq \psi \leq 90°$). Then

$$\eta = 4\varepsilon\left[\frac{v+\varepsilon\frac{\sqrt{2}}{3}}{1+\varepsilon\left(v+\varepsilon\frac{\sqrt{2}}{3}\right)}\right]^2, \tag{18}$$

and (15) becomes

$$v^3 + 3sv + 2q = 0,\quad s = -\frac{2\varepsilon^2}{9} - \frac{\xi_0}{3},\quad -q = 2^{-1/2}\left[1 + \frac{\varepsilon\xi_0}{3} + \frac{4\varepsilon^3}{27}\right], \tag{19}$$

if we put

$$\operatorname{tg}\frac{\psi}{2} = \frac{\varphi'}{|\varphi'|}\,\varepsilon^{1/2}\left(v + \varepsilon\frac{\sqrt{2}}{3}\right).$$

From the right half of (15) it follows that

$$\frac{\varphi'}{|\varphi'|} = \begin{cases} \dfrac{\left(v+\varepsilon\frac{\sqrt{2}}{3}\right)^2 - \xi_0}{\left|\left(v+\varepsilon\frac{\sqrt{2}}{3}\right)^2 - \xi_0\right|}, & 1-\varepsilon\xi_0 > 0, \\ -1 & 1-\varepsilon\xi_0 < 0, \end{cases} \tag{20}$$

and the assumption $0 \leq \psi \leq 90°$ made above imposes an additional condition on v

$$0 \leqslant \varepsilon^{1/2}\frac{\varphi'}{|\varphi'|}\left(v + \varepsilon\frac{\sqrt{2}}{3}\right) \leqslant 1, \tag{21}$$

which serves to select the necessary root from among the real solutions of (18). We note that the condition $0 \leq \psi \leq 90°$ is not obligatory. Relations (18) — (21) remain valid when ψ lies in other quadrants, too.

For real s and q, the character of the solutions of the cubic equation (18) is determined by the sign of the discriminant

$$D = q^2 + s^3 = 0{,}5 - \left(\frac{\xi_0}{3}\right)^3 + \frac{\varepsilon\xi_0}{3} + \varepsilon^2\left(\frac{2}{9} - \frac{\xi_0^2}{54}\right). \tag{22}$$

If $D > 0$, then Eq. (19) has only one real solution

$$v = (-q + D^{1/2})^{1/3} + (-q - D^{1/2})^{1/3}, \tag{23}$$

which indeed determines the maximum efficiency of the tube under consideration. If $d < 0$, all the roots of (19) are real and satisfy, within the limits of the values of ϵ and ξ_0 of interest to us, inequality (21). Inasmuch as A begins to increase from zero and decreases after reaching the first maximum, we must choose among these three solutions the one corresponding to the smallest η.

Let us consider the dependence $\eta = \eta(\xi_0)$ in the case $D \geq 0$. As follows from (18), if condition (21) is satisfied then η is a monotonic function of v, which increases together with v. It easy to see that for fixed ϵ and $\xi_0 \to \infty$ $(D > 0)$ we have

$$v \to \frac{\varepsilon\sqrt{2}}{3}, \quad \eta \to \frac{2^5\varepsilon^2}{3^2} \tag{24}$$

and η is a small quantity. When $\xi_0 = 0$ $(D > 0)$ we have

$$v \simeq 2^{1/6}\left[1 - \left(\frac{\varepsilon}{3}\right)^{2/3}\right], \tag{25}$$

and when

$$\xi_0 = \xi_{\text{opt}} \simeq \frac{3}{2^{1/3}} + 2^{1/3}\varepsilon + \frac{\varepsilon^2}{9}(2^{5/3} - 1{,}5) \tag{26}$$

$(D = 0)$, v reaches its maximum value, namely

$$v = v_{\text{opt}} = 2^{5/6}\left[1 + \frac{\varepsilon\xi_{\text{opt}}}{9}\right]. \tag{27}$$

Formulas (24), (25), and (27) show that as ξ_0 varies monotonically from $-\infty$ to approximately $3/2^{1/3}$ the efficiency of the tube increases monotonically. The values of p obtained thereby, for $\epsilon \to 0$ and equal values of ξ_0, are not more than 2 — 2.5 db greater than those given in [4], corresponding to a TWT with homogeneous beam at the entrance.

Further increase in ξ_0 leads to negative D. It is easy to see that as ξ_0 moves away from ξ_{opt} towards larger values of ξ_0, the values of v and η rapidly decrease, and approach as $\xi_0 \to \infty$ the small quantity

$$v \to -\sqrt{2}\left(\frac{\varepsilon}{3} + \frac{1}{\xi_0}\right), \quad \eta \to \frac{8\varepsilon}{\xi_0^2}. \tag{28}$$

It is interesting to compare the case of large detunings $\xi_0 \to \pm\infty$. When $\xi_0 \to -\infty$ (see (24)) the efficiency tends to a constant limit independent of ξ_0. But when $\xi_0 \to +\infty$ we have $\eta \sim \xi_0^{-2}$ and it tends asymptotically to zero.

The foregoing analysis of the relationship $\eta = \eta(\xi_0)$ shows that among all possible values of ξ_0 there is a definite region of values, within which the interaction between the electrons and the slow wave is most intense and η reaches its maximum value. Quantitatively this region can be estimated by the inequality $0 \lesssim \xi_0 \lesssim 3/2^{1/3}$.

Table 1 lists the numerical values of η for different ϵ and ξ_0. For large values of ϵ and for $\xi_0 = 3/2^{1/3}$, the tube efficiency is quite high.*

---

*A simple analysis shows that neglect of the second term of the right half of (10) in the

The upper limit of the instability region of an ordinary TWT (i.e., with homogeneous beam at the entrance) is $\xi_0 \simeq 3/4^{1/3}$ and is found to be lower than in the tube which we are considering. We see therefore that the width of the instability region of traveling wave tubes depends on the excitation method.

Table 1

Values of η as functions of ξ_0 and ϵ

ξ_0	ε				
	$\varepsilon \to 0$	0.05	0.1	0.15	0.2
0	$5.04\,\varepsilon$	0.21	0.36	0.48	0.58
0.75	$7.21\,\varepsilon$	0.31	0.54	0.70	0.82
$3/4^{1/3}$	$11.0\,\varepsilon$	0.44	0.70	0.86	0.95
$3/2^{1/3}$	12.7ε	0.48	0.76	0.90	0.98
$4/2^{1/3}$	$0.9233\,\varepsilon$	0.046	0.093	0.140	0.188

The optimal length of the tube $\zeta = \zeta_m$, at which maximum η is attained, is given by the equation

$$\eta' = \pm\sqrt{8\varepsilon\eta - \eta^2\left(\frac{1}{\varepsilon}\,\frac{1-\sqrt{1-\eta}}{1+\sqrt{1-\eta}} - \xi_0\right)^2}, \tag{29}$$

which follows from (14) and (15); ζ_m is given by a linear combination of complete elliptic integrals of the first and third kinds:

$$\zeta_m = L\int_0^{\pi/2}\left[1 - \frac{2\varepsilon}{1+\varepsilon\xi_0}\left(\alpha_1 + \frac{\beta_1-\alpha_1}{1-\frac{\beta_1^2}{\alpha_1^2}\cos^2\varphi}\right)\right]\frac{d\varphi}{\sqrt{1-\frac{k}{1+k}\sin^2\varphi}}, \tag{30}$$

where

$$L = \frac{2(\beta_1-\alpha_1)}{\sqrt{\alpha_1\left(\alpha_1 - u\frac{1+\varepsilon\xi_0}{2}\right)\left[\beta_1^2+\beta_1\left(\frac{1+\varepsilon\xi_0}{2}u-2\xi_0\right)+4u^{-1}\right](1+k)}};$$

$$k = \frac{\beta_1^2}{\alpha_1^2}\,\frac{\alpha_1^2+\alpha_1\left(\frac{1+\varepsilon\xi_0}{2}u-2\xi_0\right)+4u^{-1}}{\beta_1^2+\beta_1\left(\frac{1+\varepsilon\xi_0}{2}u-2\xi_0\right)+4u^{-1}}; \quad u = \frac{2\left(v+\varepsilon\frac{\sqrt{2}}{3}\right)^2}{1+\varepsilon\left(v+\varepsilon\frac{\sqrt{2}}{3}\right)^2};$$

$$\beta_1 = -\alpha_1 - \frac{4u^{-1}}{\frac{1+\varepsilon\xi_0}{2}u-\xi_0} =$$

$$= \sqrt{\left(\frac{2u^{-1}}{\frac{1+\varepsilon\xi_0}{2}u-\xi_0}\right)^2 + \frac{1+\varepsilon\xi_0}{\frac{1+\varepsilon\xi_0}{2}u-\xi_0}} - \frac{2u^{-1}}{\frac{1+\varepsilon\xi_0}{2}u-\xi_0};$$

v is the necessary root of Eq. (19). Since there are no published tables of complete elliptic integrals of the third kind, it is necessary to expand $\left(1-\frac{\beta_1^2}{\alpha_1^2}\cos^2\varphi\right)^{-1}$ in (30) in powers of $\frac{\beta_1^2}{\alpha_1^2}\cos^2\varphi$, using the smallness of $\beta_1{}^2/\alpha_1{}^2$, and retain the first two terms of the expansion. Then ζ_m is given in terms of the tabulated elliptic integrals of the first and second kind (see, for example [5]) and can be readily calculated.

---

equation of motion is valid when $\eta \leq 95\%$. When $\eta < 95\%$ the power flux corresponding to the dimensionless force $-\epsilon^3 w''/8$ increases and, for example, when $\xi_0 = 3 \times 2^{-1/3}$ and $\epsilon = 0.2\,(\eta = 0.98)$, it decreases η by 2%.

The calculations performed have shown that when ξ_0 lies in the interval $0 \lesssim \xi_0 \lesssim 3/4^{1/3}$, the optimum length of the tube ζ_m changes slowly (and almost linearly) together with ϵ, and with accuracy of about 10% it is equal to $\zeta_m = 2$ for $\epsilon \to 0$ and $\zeta_m = 1.3$ for $\epsilon = 0.2$. When $\xi = 3/2^{1/3}$, the optimum length is $\zeta_m = 4.5$ for $\epsilon \to 0$ and $\zeta_m = 2.2$ for $\epsilon = 0.2$.

These results show that the effective length of the instrument considered is approximately 1/3 as large as the effective length of an ordinary TWT with homogeneous beam at the input. The noticeable (almost two-fold) increase in ζ_m as $\xi_0 \to \xi_{opt}$ is connected with the fact that as $\xi_0 \to \xi_{opt}$ only the degree of transfer of energy from the electrons to the field is increased, but not the rate (i.e., the gain per unit length) of this transfer. An analogous effect occurs also in an ordinary TWT with homogeneous beam at the entrance.

3. INCREASE OF EFFICIENCY OF THE TUBE

The design of a TWT tube with large coupling coefficient is technically a complicated problem. Therefore, a search for methods of increasing efficiency of tubes with low coupling coefficients ($\epsilon \lesssim 0.05$) is quite urgent. In Eqs. (1) — (3) are contained two parameters h_0 and R_0, which determine the phase velocity of the slow wave in the "cold" (i.e., when $j_0(t_0) = 0$) system and the degree of its interaction with the electron beam. These quantities were assumed constant in the foregoing discussions. The question is, can we increase the tube efficiency by smoothly varying h_0 or R_0 along the tube?

The value of R_0 is very difficult to change within a wide range. We shall, therefore, consider below only the effect of the variable h_0 on the efficiency η of the tube. The problem of the optimum choice of $h_0 = h_0(\zeta)$, at which η reaches itsmaximum, is quite rigorously formulated. By supplementing the system (1)-(3) with the energy integral contained in it and varying the latter with respect to h_0 with fulfillment of conditions (1)-(3), we can obtain Euler's equation for the extremal function $h_0 = h_0(\zeta)$, and certain additional conditions, which relate the boundary values of the functions $E(z)$, $h_0(z)$, and $t(z)$ with each other and which determine the position z_0 of the extremum point of the function $|E(z)|$.

The problem so formulated is quite complicated, since its solution must satisfy definite constraints imposed on $E(z)$, $h_0(z)$, $t(z)$ and their derivatives at the point $z = 0$ and $z = z_1$, of which z_1 (the extremum point) must be determined during the process of the solution. A simple estimate of the extent to which such a problem is laborious shows that it is easier to specify $h_0(\zeta)$ in the form of an arbitrary function, which depends on several parameters, and the best choice of the latter is made with the aid of trial solutions, as was done in [6]. Nonetheless, a rigorous formulation of the problem of finding the optimum $h_0(\zeta)$ leads to interesting qualitative conclusions concerning the extremal properties of the investigated interaction process between ideally focused bunches and the self-radiation field (diploma thesis of Yu. A. Khristev).

It is found first of all that it is possible to attain an absolute field maximum $(\partial t(z_1)/\partial z = \infty)$ only with the aid of a special variation $R_0 = R_0(z)$. Since R_0 was assumed constant in [6], the extrema obtained there are not the extrema of the given type. The second group of extrema is characterized by the satisfaction of the condition $(w(\zeta_1) - \varphi(\zeta_1)) = 0$, in which the acceleration of the bunch vanishes, and the field of the slow wave does not act on the bunch as a whole at the point of extremum. A remarkable property of extrema of these kind is that to realize it one does not need a unique specification of the boundary values of $E(z)$, $t(z)$ and $h_0(z)$ and of their derivative. We note that in [6] this fact manifests itself, apparently, in that $|E(z_1)|$ was practically independent of ξ_0 under a suitable choice of the parameter α introduced in [6].

There exists, finally, a third group of additional conditions, which lead to the satisfaction of the necessary condition of extremum $A'(\zeta_1) = 0$. It is characterized by a unique specification of the boundary values and leads to $w''(\zeta_1) \neq 0$. This means that at the point $z = z_1$ the field either already slows down or already accelerates the electron bunch, and is not equal to its maximum value. Therefore, these conditions were not investigated in detail, since they have no physical interest for our problem.

4. INVESTIGATION OF ISOCHRONOUS MODE

A tube with variable $h_0 = h_0(\zeta)$ can be called "isochronous," since a definite phase shift between the bunch and the field is maintained in it by artificially slowing down the cold wave ("synchronization").

In the case of variable $h_0 = h_0(\zeta)$, there naturally appear in the fundamental equations two wave numbers

$$h_0 = h_0(\zeta) = \frac{\omega}{v_0}(1 + \varepsilon\xi(\zeta));\ h_0(0) = h_{00} = \frac{\omega}{v_0}(1 + \varepsilon\xi_0) \tag{31}$$

and R_0 is found to be a function of ζ. The dimensionless equations for the isochronous tube follow directly from Eqs. (80.09) and (80.04) of [7], which represent the field acting on the electrons. Taking h_0, R_0 in (5) to mean h_{00} and $R_0(0)$, we obtain in lieu of (8) and (9)

$$w'' = -(1 + \varepsilon\xi_0 + \varepsilon w')^3 \mathrm{Re}\left[F_+ e^{i\int_0^z h_0(\zeta)dz} + F_- e^{-i\int_0^z h_0(\zeta)dz}\right], \tag{32}$$

$$F'_\pm = \mp 2\frac{R_0(\zeta)}{R_0(0)} e^{i\left[h_{00}z \mp \int_0^z h_0(\zeta)dz + w\right]} \tag{33}$$

The dimensionless power is defined here by the formula

$$p = \frac{1}{2}\int_0^\zeta \frac{R_0(0)}{R_0(\zeta)} d|F_+|^2. \tag{34}$$

In the case of a cylindrical delay line (see [3]) the value of $R_0(\zeta)$ in the first approximation in powers of ϵ is

$$\frac{R_0(0)}{R_0(\zeta)} = 1 + \varepsilon\gamma[\xi(\zeta) - \xi_0],\ \gamma = \frac{1}{2}\frac{\left(\frac{\omega b}{v_0}\right)^2}{1 - \frac{1}{8}\left(\frac{\omega b}{v_0}\right)^2}, \tag{35}$$

where b is the angle of the electron beam. According to the experimental results given in [1] (p. 94), $\omega b/v_0$ must not be assigned values greater than 0.8, for when $\omega b/v_0 > 0.8$ the efficiency of the tube is reduced by the stratification of the beam. Corresponding to these values of $\omega b/v_0$ are $0 \leq \gamma \lesssim 1/3$.

Proceeding now from Eqs. (32) and (33) to their first integrals and using for F_+ expression (12), we obtain

$$\varphi' A = \pm\sqrt{\left(2\frac{R_0(\zeta)}{R_0(0)}\right)^2 - A'^2} = \frac{R_0(\zeta)}{AR_0(0)}\int_0^A (\xi_0 + w' - \xi)\frac{R_0(0)}{R_0(\zeta)} A\,dA, \tag{36}$$

which enable us to calculate directly the maximum values of the amplitude A. For this purpose we need only put in (36) $A' = 0$ and integrate there.

As before, w' is determined with the aid of (13), where the value of p is taken from (34) and $\xi(\zeta)$ must be specified in explicit form. Since the slowing down of the electrons in the tube is equivalent to an increase in the efficiency η, it is meaningful to assume $\xi(\zeta)$ to be an explicit function of η or, what is the same, an explicit function of A^2. The latter representation

$$\xi(\zeta) = \xi_0 + \sum_{k=1}^{\infty} q_k \varepsilon^{k-1} A^{2k} \tag{37}$$

appears to us to be more expedient.

We have already noted above that the use of isochronism to increase the tube efficiency is meaningful only for small ϵ, when the efficiency of the homogeneous tube is small. Therefore, assuming ϵ in (36) to be a small quantity and expanding the right half of (36) in this quantity, we obtain the following equation for the maximum field amplitude.

$$\pm 2 = -\frac{\xi_0}{2} A + \frac{A^3}{4}\left[\frac{1}{4} - q_1(1+\varepsilon\gamma\xi_0)\right] + \varepsilon\frac{A^5}{6}\left[\frac{3}{32} - q_2(1+\varepsilon\gamma\xi_0) + \right.$$
$$\left. + \gamma q_1\left(\frac{3}{8} - q_1\right)\right] + \varepsilon^2\frac{A^7}{8}\left[\frac{1}{16} - q_3(1+\varepsilon\gamma\xi_0) + \frac{\gamma q_2}{3} + \right.$$
$$\left. + \gamma q_1\left(\frac{3}{16} + \frac{\gamma q_1}{8} - 2q_2\right)\right] + \dots \tag{38}$$

For comparison we first consider the solution of (38) in the nonisochronous case $q_k = 0$. Putting in the limit $\epsilon = 0$, we obtain a cubic equation which has a single real root when $\xi_0 < 3.2^{-1/3}$:

$$A = A_0(\xi_0) = 2^{4/3}\left[\sqrt[3]{1+\sqrt{1-2\left(\frac{\xi_0}{3}\right)^3}} + \sqrt[3]{1-\sqrt{1-2\left(\frac{\xi_0}{3}\right)^3}}\right], \tag{39}$$

which reaches its maximum when $\xi_0 = 3.2^{-1/3}$, and when the other two roots of (38) become equal. We note that (38) enables us to calculate small additions to $A_0(\xi_0)$, corresponding to finite values of ϵ.

In the simplest isochronous case $q_k = q_1\delta_{k,1}$ ($\delta_{k,p} = 0$, if $k \neq p$, and $\delta_{k,p} = 1$ if $k = p$), Eq. (38) again is cubic when $\epsilon = 0$. Choosing q_1 in this equation

$$q_1 = \frac{1}{4} - \frac{1}{2}\left(\frac{\xi_0}{3}\right)^3. \tag{40}$$

such that for each ξ_0 the two roots of (38) remain equal, we obtain

$$A = A_1(\xi_0) = \frac{12}{|\xi_0|}. \tag{41}$$

It is easy to show that the Eq. (40) for q_1 indeed leads to extremal values of $A_1(\xi_0)$ (with respect to q_1.

In the analysis of (38) we undertook to obtain for $\epsilon = 0$ larger values of $A_1(\xi_0)$ (than given by (41)) by retaining in (38) terms of higher order with respect to $A(q1 = q'_1\epsilon^{-1}$, $q2 = q'_2\epsilon^{-2}, \dots)$. However, the results were unfavorable. This means that a quadratic variation of ξ with $A(q_k = q1\ \delta_{k,1})$ is the most expedient for slowing down the wave in the case of ideal focusing of the bunches and infinitesimally small values of ϵ.

As can be seen from (41), $A_1(\xi_0)$ reaches its maximum when $\xi_0 \to 0$. The infinite value of $A_1(0)$ is the result of the idealization $\epsilon = 0$. The correct value of $A_1(0)$ can be obtained directly from (38) by putting there $\xi_0 = 0$, $q_k = \frac{1}{4}\ \delta_{k,1}$ (see (40)). We then get

$$A_1(0) \simeq 2\left(\frac{4\varepsilon^{-1}}{1+\frac{\gamma}{3}}\right)^{1/5}\left[1 + 0{,}7\varepsilon^{3/5}\frac{1+\frac{3\gamma}{4}+\frac{\gamma^2}{8}}{1+\frac{7\gamma}{15}}\right]^{-1}. \tag{42}$$

The second term in the square brackets of (42) characterizes the accuracy of calculation of $A_1(0)$. We see readily that when $0 \leq \gamma \lesssim 1/3$ the amplitude $A_1(0)$ is practically independent of γ.

If we now choose the coefficients ξ_0, q_1, and q_2 in such a way that the first three terms ($\xi_0 = 0$, $q_k = \frac{1}{4}\ \delta_{k,1} + \frac{3+\gamma}{32}\ \delta_{k,2}$) vanish in the right half of (38), we find that

$$A_1(0) \simeq 2\left(\frac{2\varepsilon^{-2}}{1+\frac{\gamma}{2}+\frac{\gamma^2}{24}}\right)^{1/7}. \tag{43}$$

When $\epsilon < 0.125$ this value of $A_1(0)$ exceeds the value given by (42). This means that in a real tube ($\epsilon \neq 0$) the efficiency of different synchronization methods at $\xi_0 = 0$ depends in a complicated manner on the value of ϵ. It is easy to see that the right half of (43) also depends little on γ in the interval $0 \leq \gamma \lesssim 1/3$.

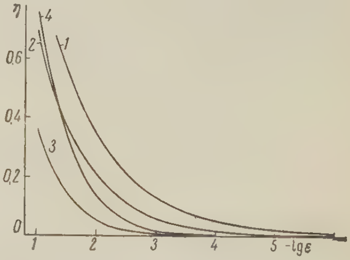


Fig. 1. Dependence of η on ϵ when $\gamma = 0$.

1 — Isochronous mode for $\xi_0 = 0$, $q_k = \frac{1}{4}\,\delta_{k,1} + \frac{3}{32}\,\delta_{k,2}$ (see (43)); 2 — isochronous mode for $\xi_0 = 0$, $q_k = \frac{1}{4}\,\delta_{k,1}$ (see (42)); 3 — non-isochronous mode for $\xi_0 = 0$ (see (39)); 4 — non-isochronous mode for $\xi_0 = 3.2^{-1/3}$ (see (39)).

Fig. 1 shows the dependence of η on ϵ for $\xi_0 = 0$ in the isochronous mode (curves 1,2). For comparison, Fig. 1 shows also the lines $\eta = \eta(\epsilon)$, corresponding to the non-isochronous ($q_k = 0$) mode with $\xi_0 = 0$ (curve 3) and $\xi_0 = 3.2^{-1/3}$ (curve 4). The comparison shows that while isochronism increases the efficiency of the tube at a fixed value of ξ_0, the limiting efficiencies of the homogeneous tube ($q_k = 0$, $\xi_0 = 3.2^{-1/3}$) are found to be lower than the optimum isochronous efficiencies ($\xi_0 = 0$) only when $\epsilon \lesssim 0.05$. It is quite interesting to compare the values of A given by (41) with the calculations carried out for other methods of excitation of the delay line and $\epsilon = 0$ (see Table 2). It is convenient to take $\xi_0 = 0.75$, for as $\xi_0 \to 0$ we cannot use (41).

Table 2

Values of A as $\epsilon \to 0$, $\xi_0 = 0.75$, and different q_k and $j_0(t_0)$

№	A	$j_0(t_0)$	q_k
1	2.9	j_0	$0.75\delta_{k.0}$
2	3.8	$j_0\delta\left(\frac{t_0}{T}\right)$	$0.75\delta_{k.0}$
3	3.8	$j_0\frac{10}{2}\left[\frac{t_0}{\lvert t_0\rvert} + \frac{0.1\cdot 2\pi - \omega t}{\lvert 0.1\cdot 2\pi - \omega t\rvert}\right]$	$0.75\delta_{k.0}$
4	3.8	j_0	$0.75\delta_{k.0} + \frac{0.833}{4}\,\delta_{k.2}$
5	16	$j_0\delta\left(\frac{t_0}{T}\right)$	$0.75\delta_{k.0} + \frac{0.969}{4}\,\delta_{k.2}$
6	5.75	$j_0\,\frac{10}{2}\left[\frac{t_0}{\lvert t_0\rvert} + \frac{0.1.2\pi - \omega t_0}{\lvert 0.1\cdot 2\pi - \omega t_0\rvert}\right]$	$0.75\delta_{k.0} + \frac{0.750}{4}\,\delta_{k.2}$
7	7.8	$j_0\,\frac{10}{2}\left[\frac{t_0}{\lvert t_0\rvert} + \frac{0.1\cdot 2\pi - \omega t_0}{\lvert 0.1\cdot 2\pi - \omega t_0\rvert}\right]$	$0.75\delta_{k.0} + \frac{0.500}{4}\,\delta_{k.2} + \frac{0.050}{16}\,\delta_{k.4}$

The first line of Table 2 yields the value of A_{max} obtained in [4] for excitation of a homogeneous helix ($q_k = \xi_0\,\delta_{k,0}$) by a homogeneous beam of electrons ($j_0(t_0) = j_0$). The second gives A_{max} for excitation of a homogeneous helix by ideally focused bunches. The result is somewhat higher. In the third line is given the value of A_{max} obtained when the delay line is excited by a periodic sequence of bunches with duration of 0.1 cycle (equal to 2π in dimensionless units). We see that A_{max} coincides with the preceding result. In the fourth line is given the same result obtained in the simplest synchronization method ($q_k = \xi_0\,\delta_{k,0} + q_2\,\delta_{k,2}$) when the electrons and the wave move in a tube excited by a continuous flow of electrons at the entrance (see [6]), while the fifth and sixth lines give the result when the line is excited by ideally focused and extensive bunches, respectively. As can be seen, the results are highly divergent in the isochronous mode, in distinction from the non-isochronous mode (lines 1, 2, 3).

The seventh line of Table 2 shows that the efficiency of the tube can be appreciably increased (doubled in terms of power) even at small spreading of the exciting bunches, by retaining in (39) terms with higher order of A. This difference from ideally-focused bunches is probably due to the fact that the optimum synchronization of the motion of the wave in extensive bunches includes not only a sufficiently rapid slowing down of the "cold" wave, but also a concentration of the initial bunch, whereas the "ideally focused" bunches (four) can no longer be concentrated. Fig. 2 shows the function $P = p(\zeta)$ for the seventh line.

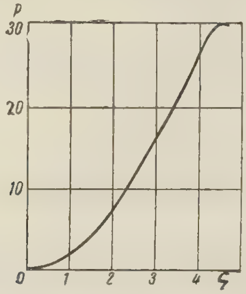


Fig. 2. Dependence of p on ζ for i_0

$$(t_0)=i_0\frac{10}{2}\left[\frac{t_0}{|t_0|}+\frac{0.1\cdot 2\pi-\omega t_0}{|0.1\cdot 2\pi-\omega t_0|}\right]$$

and $q_k = 0.75\,\delta_{k,0} + 0.125\,\delta_{k,2} + \frac{0.05}{16}\,\delta_k.$

Comparison of the values of q_2 in the fourth to sixth lines of Table 2 is evidence that the rate of transfer of energy from the electrons to the wave and vice versa depends appreciably on the method of excitation of the traveling wave tube.

REFERENCES

1. Traveling wave tube. Collection of translations, edited by V.T. Ovcharov, GEI, 1959.
2. L.A.Vaynshteyn. Radiotekhnika i elektronika, 1957, 2, 6, 688; 1957, 2, 7, 883.
3. L.A.Vaynshteyn, ZhTF. Journal of Technical Physics. 1956, 26, 1, 1-41.
4. L.A. Vaynshteyn. Radiotekhnika i elektronika, 1957, 2, 8, 1027.
5. E. Jahnke and F. Emde. Tables of Functions, Dover, 1943.
6. G.F. Filimonov. Radiotekhnika i elektronika, 1953, 3, 1, 85.
7. L.A. Vaynshteyn. Electromagnetic Waves, Soviet Radio Press, 1957

Received by editor 19 November 1960

GENERAL DISPERSION EQUATION OF A TRAVELING WAVE TUBE WITH PERIODIC SLOW-WAVE SYSTEM

V.N. Shevchik, A.A. Kurayev

On the basis of premises concerning the cascade interaction between electrons and the field of a periodic slow-wave system, we obtain the dispersion equation of a TWT in the most general case for the one-dimensional problem. We indicate the boundary of applicability of the direct harmonic analysis.

INTRODUCTION

Modern high-power traveling-wave tubes employ as a rule periodic slow-wave systems (such as coupled resonators, periodically-loaded waveguides, etc.). Inhomogeneous slow-wave systems (for example, combs, facing rods, etc.) are also used in the millimeter band.

The use of inhomogeneous lines as slow-wave systems for amplifiers makes it necessary to account for, in the analysis of such amplifiers, the discrete nature of the interaction between the electrons and the field of the line. In harmonic analysis, the field of the line is expanded in a Fourier series in the spatial harmonics, and the interaction between the electron beam with each of the harmonics is calculated [1 — 4]. The resultant field of the line is represented in the form of a sum of the terms of the infinite series; the magnitude of each term depends on the degree of interaction between the electrons and the field of a given harmonic:*

$$E = \frac{1}{2}(\Gamma^2 + \beta_0^2) i \sum_n \frac{\left(\frac{E^2}{\beta^2 P}\right)_n \Gamma_n^3}{(\Gamma_n^2 + \beta_0^2)(\Gamma_n^2 - \Gamma^2)} \cdot \tag{I}$$

In the case when the electrons are synchronized with one of the spatial harmonics, the term of the series describing the interaction with this space harmonic becomes considerably greater than all the other terms. In this case the expression for E simplifies to

$$E = \left[\frac{\Gamma^2 \Gamma_1 \left(\frac{F^2}{\beta^2 P}\right)}{2(\Gamma_1^2 - \Gamma^2)} - \frac{i \Gamma^2}{\omega C_1}\right] i, \tag{II}$$

and this leads subsequently to a simplification of the characteristic equation of the tube and the possibility of reducing it to a third-degree equation. However, when the electrons are simultaneously synchronized with two spatial harmonics (as takes place on the limits of the pass band in inhomogeneous systems), harmonic analysis leads to cumbersome expressions, which cannot be simplified by introducing convenient variables. At the same time, analysis of the operation of the amplifier at the boundaries of the passband and beyond these boundaries is of interest [5, 6, 16].

There exists another method of analysis, which takes into account the interaction of the electrons with all the harmonics simultaneously. It is based on the notion of cascade bunching of electrons and the use for analysis of an amplifier with slow-wave systems such that their period can be broken up with sufficient accuracy into two parts, one free of fields (drift space) and one in which the field is constant (interaction space). Both spaces can be arbitrary in length, depending on the type of the slow-wave system.

---

*Formulas (I) and (II) are cited in accordance with [1].

The cascade analysis procedure has been developed in [7 — 9]. This analysis is used in [6, 10, 11, 16] to obtain a self-consistent solution in some particular cases. The present paper contains a further development of this method and its application to the analysis of a TWT in the most general case, for the one-dimensional problem, particularly with account taken of space charge, line losses, and finite transit angle in interaction space.

The analysis is carried out for the case of small signals. The condition of smallness of the signal (and of applicability of the elementary electron-wave theory used in the present paper) can be represented in the form

$$\frac{V_{n-1}}{2V_0}\,\frac{\omega_p}{\omega_p} \ll 1\,.$$

Relativistic effects are disregarded. Nor do we consider the stopping of electrons in the system and the associated decrease in the average current along the length of the tube.

BASIC SYMBOLS

θ_1 — transfer constant of cell of equivalent line without account of induced current.
θ — transfer constant of cell equivalent line with account of interaction between the line field and the electron beam.
$\varphi^{(2)}$ — transit angle of electrons in drift space.
$\varphi^{(1)}$ — transit angle of electrons in interaction space.
φ — transit angle of electrons through the period of the slow-wave system.
E_2 — field intensity of the ac component of the space charge.
ω_p — $\sqrt{4\pi\eta\rho_0}$ — plasma frequency.
R_c — coupling resistance.
$G = \sqrt[3]{I_0R_c/4V_0}$ — amplification parameter.
l — length of period of slow-wave system.
d — length of interaction drift space.
The remaining symbols are explained in the text.

1. SOLUTION OF ELECTRONIC EQUATIONS

General equations. We consider the interaction space (Fig. 1). The connection between the beam parameters (ρ, v, E_2) and their connection to the applied field of intensity E_1 is given by the system of equations*

$$\begin{gathered}\rho = \frac{1}{4\pi}\frac{\partial E_2}{\partial x}\,,\\ \frac{d\tilde{v}}{dt} = \eta\,(E_1 + E_2),\\ \frac{\partial(\rho\tilde{v})}{\partial x} = -\frac{\partial\rho}{\partial t}\,.\end{gathered} \tag{1}$$

From the first and third equations of the system we readily obtain

$$\frac{dE_2}{dt} = -4\pi\rho\tilde{v}$$

or in the linear approximation

$$\frac{dE_2}{dt} = -4\pi\rho_0\tilde{v}.$$

---

*The system (1) is presented in a coordinate frame moving with the average electron velocity v_0.

Using the second equation, we obtain

$$\frac{d^2\tilde{v}}{d\tau^2} + 4\pi\rho_0\tilde{v} = j\omega\eta E_1 e^{j\omega(t_1+\tau)}.$$

Introducing the notation $\varphi = \omega\tau$, $\mu = \eta E/\omega v_0$, $a = \omega_p/\omega$, we obtain an equation for the time component of the velocity in the interaction space

$$\frac{d^2\tilde{v}}{d\varphi^2} + a^2\tilde{v} = jv_0\mu e^{j(\omega t_1+\varphi)}. \tag{2}$$

For the drift space, where $E_1 = 0$, we obtain

$$\frac{d^2\tilde{v}}{d\varphi^2} + a^2\tilde{v} = 0. \tag{3}$$

The general solution of (2) has the form

$$\tilde{v} = C^{(1)}e^{ja\varphi^{(1)}} + C^{(2)}e^{-ja\varphi^{(1)}} - \frac{jv_0\mu}{1-a^2}e^{j(\omega t_1+\varphi)}. \tag{4}$$

The general solution of (3) is

$$\tilde{v} = A^{(1)}e^{ja\varphi^{(2)}} + A^{(2)}e^{-ja\varphi^{(2)}} \tag{5}$$

The arbitrary constants must be determined from the initial condition ($\varphi = 0$). For this we must specify $\tilde{v}$ and $d\tilde{v}/dt$ at the start of the space. The derivative $d\tilde{v}/dt$ is connected with E_2 through the equation of motion, i.e., specification of $d\tilde{v}/dt$ is tantamount to specifying E_2. Using the equation of motion and the additional relation

$$\frac{d\tilde{i}}{dt} = \rho_0\frac{d\tilde{v}}{\partial t},$$

we obtain an expression for E_2 and $\tilde{i}$ in terms of the arbitrary constants C_1, $C_2(A_1, A_2)$.

For the modulation space we get

$$\begin{gathered} E_2 = j\frac{\omega}{\eta}a\left(C^{(1)}e^{ja\varphi^{(1)}} - C^{(2)}e^{-ja\varphi^{(1)}}\right) + \frac{v_0}{\eta}\mu\frac{\omega a}{1-a^2}e^{j(\omega t_1+\varphi)}, \\ \tilde{i} = \frac{\rho_0}{a}\left(C^{(1)}e^{ja\varphi^{(1)}} - C^{(2)}e^{-ja\varphi^{(1)}}\right) - j\frac{\rho_0 v_0\mu}{1-a^2}e^{j(\omega t_1+\varphi)}. \end{gathered} \tag{6}$$

Analogously for the drift space

$$\begin{gathered} E_2 = j\frac{\omega}{\eta}a\left(A^{(1)}e^{ja\varphi^{(2)}} - A^{(2)}e^{-ja\varphi^{(2)}}\right). \\ \tilde{i} = \frac{\rho_0}{a}\left(A^{(1)}e^{ja\varphi^{(2)}} - A^{(2)}e^{-ja\varphi^{(2)}}\right). \end{gathered} \tag{7}$$

We now consider the complete diagram of the interaction stage (Fig. 1). We see from the diagram that the determination of $\tilde{v}_k$, E_{2k} and $\tilde{i}_k$ reduces to a successive substitution of the initial conditions ($\tilde{v}_n/\varphi_n = 0$ and $E_{2n}/\varphi_n = 0$ or $\tilde{v}_n/\varphi_n = 0$ and $\tilde{i}_n/\varphi_n = 0$) into the corresponding expressions for $\tilde{v}_n$ and E_{2n} or $\tilde{v}_n$ and $\tilde{i}_n$ for the determination of $C_n^{(1)}$ and $C_n^{(2)}$ ($A_n^{(1)}$ and $A_n^{(2)}$), which is equivalent to joining the solutions of Eqs. (2) and (3) on the boundaries of neighboring spaces. Let us write these solutions for the k-th slot, introducing μ_k and μ_1 in accordance with the equation $\mu_k = \mu_1 e^{-(k-1)\theta}$:

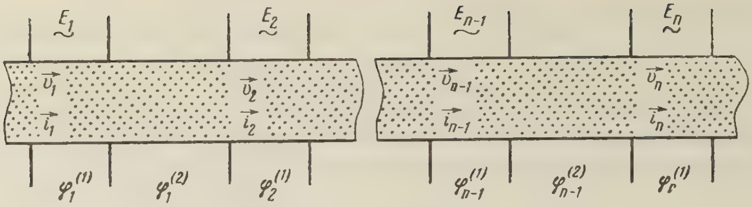


Fig. 1

$$
\begin{aligned}
\tilde{v}_k &= C_k^{(1)} e^{ja\varphi_k^{(1)}} + C_k^{(2)} e^{-ja\varphi_k^{(1)}} - \frac{j\mu_1 v_0}{1-a^2} e^{j\omega t_1} e^{j\left(\varphi_k^{(1)} + \sum\limits_m^{k-1} \varphi_m\right) - (k-1)\theta} \\
E_{2k} &= j\frac{\omega}{\eta} a \left(C_k^{(1)} e^{ja\varphi_k^{(1)}} - C_k^{(2)} e^{-ja\varphi_k^{(1)}}\right) + \frac{v_0}{\eta}\mu \frac{\omega a}{1-a^2} e^{j\omega t_1} e^{j\left(\varphi_k^{(1)} + \sum\limits_m^{k-1} \varphi_m\right) - (k-1)\theta} \qquad (8) \\
\tilde{i}_k &= \frac{\rho_0}{a}\left(C_k^{(1)} e^{ja\varphi_k^{(1)}} - C_k^{(2)} e^{-ja\varphi_k^{(1)}}\right) - j\frac{\rho_0 v_0 \mu}{1-a^2} e^{j\omega t_1} e^{j\left(\varphi_k^{(1)} + \sum\limits_m^{k-1} \varphi_m\right) - (k-1)\theta}
\end{aligned}
$$

For the k-th drift space we have

$$
\begin{aligned}
\tilde{v}_k &= A_k^{(1)} e^{ja\varphi_k^{(2)}} + A_k^{(2)} e^{-ja\varphi_k^{(2)}}, \\
E_{2k} &= j\frac{\omega}{\eta} a \left(A_k^{(1)} e^{ja\varphi_k^{(2)}} - A_k^{(2)} e^{-ja\varphi_k^{(2)}}\right), \qquad (9) \\
\tilde{i}_k &= \frac{\rho_0}{a}\left(A_k^{(1)} e^{ja\varphi_k^{(2)}} - A_k^{(2)} e^{-ja\varphi_k^{(2)}}\right).
\end{aligned}
$$

The initial conditions and arbitrary constants for the k-th slot. Let us solve the problem subject to the following initial conditions:

$$\tilde{v} = 0, \quad E_2 = 0 \ (\tilde{i} = 0) \ \text{for} \ \varphi_1^{(1)} = 0, \qquad (10)$$

i.e., the beam at the entrance to the system will be assumed unmodulated. Under these initial conditions we obtain from (8)

$$
\begin{aligned}
C_1^{(1)} &= \frac{jv_0\mu_1}{2(1-a)} e^{j\omega t_1}, \\
C_1^{(2)} &= \frac{jv_0\mu_1}{2(1+a)} e^{j\omega t_1}. \qquad (11)
\end{aligned}
$$

For the k-th slot, we obtain recursion formulas for the coefficients $C_k^{(1)}$ and $C_k^{(2)}$:

$$
\begin{aligned}
C_k^{(1)} &= \frac{jv_0\mu_1}{2(1-a)} e^{j\left(\omega t_1 + \sum\limits_m^{k-1} \varphi_m\right)} e^{-(k-1)\theta} \left(1 - e^{\theta - j\varphi_{k-1} + j\varphi_{k-1}^{(1)} + ja\varphi_{k-1}^{(2)}}\right) + C_{k-1}^{(1)} e^{ja\varphi_{k-1}}. \\
& \qquad (12) \\
C_k^{(2)} &= \frac{jv_0\mu_1}{2(1+a)} e^{j\left(\omega t_1 + \sum\limits_m^{k-1} \varphi_m\right)} e^{-(k-1)\theta} \left(1 - e^{\theta - j\varphi_{k-1} - j\varphi_{k-1}^{(1)} - ja\varphi_{k-1}^{(2)}}\right) + C_{k-1}^{(2)} e^{-ja\varphi_{k-1}}.
\end{aligned}
$$

To simplify (12), and also to carry out succesive transformations, we used the smallness of the signal $\mu_1 \ll 1$.

Let us consider the function $f = \sum\limits_m^{k-1} \varphi_m = (k-1)\varphi_0 + \Delta\varphi(\mu_1)$, wherein $\Delta\varphi(\mu_1) = 0$ at the point $\mu_1 = 0$.

We expand this function in a Maclaurin series in μ_1:

$$f(\mu_1) = (k-1)\varphi_0 + \mu_1 \frac{\partial f(\mu_1)}{\partial \mu_1}\bigg|_{\mu_1=0} + \dots .$$

We note that $\partial f(\mu_1)/\partial\mu_1 \to \infty$ at the overtaking points. However, all the foregoing calculations are within the framework of the electron-wave theory, in which overtaking cannot be considered. The condition $\frac{V_{n-1}}{2V_0}\frac{\omega}{\omega_p} \ll 1$ serves as a guarantee that no overtaking takes place. Thus, within the framework of the linear theory we can write

$$\sum_{m}^{k-1} \varphi_m \simeq (k-1)\varphi_0. \tag{13}$$

Taking (13) into account, the expressions of (12) assume the form

$$\begin{aligned} C_k^{(1)} &= \frac{jv_0\mu_1}{2(1-a)} e^{j[\omega t_1-(k-1)T]} e^{(k-1)\alpha}\left(1 - e^{-(\alpha-jT)+j\varphi_0^{(1)}+ja_0^{(2)}}\right) + C_{k-1}^{(1)} e^{ja\varphi_0}, \\ C_k^{(2)} &= \frac{jv_0\mu_1}{2(1+a)} e^{j[\omega t_1-(k-1)T]} e^{(k-1)\alpha}\left(1 - e^{-(\alpha-jT)+j\varphi_0^{(1)}-ja\varphi_0^{(2)}}\right) + C_{k-1}^{(2)} e^{-ja\varphi_0}, \end{aligned} \tag{14}$$

where the following notation is used

$$\theta = -\alpha + j\theta',\ T = \theta' - \varphi_0,\ \varphi_0 = \varphi_0^{(1)} + \varphi_0^{(2)}$$

— unperturbed transit angle through one period of the slow-wave system.

Using (14) and (11), we can represent the expressions for $C_k^{(1)}$ and $C_k^{(2)}$ in the form of the sums

$$\begin{aligned} C_k^{(1)} &= C_1^{(1)}\left(L_1 e^{(k-1)(\alpha-jT)} + L_1 e^{(k-2)(\alpha-jT)} e^{ja\varphi_0} + \right. \\ &\quad \left. + L_1 e^{(k-3)(\alpha-jT)} e^{j2a\varphi_0} + \dots e^{(k-1)ja\varphi_0}\right), \\ C_k^{(2)} &= C_1^{(2)}\left(L_2 e^{(k-1)(\alpha-jT)} + L_2 e^{(k-2)(\alpha-jT)} e^{-ja\varphi_0} + \right. \\ &\quad \left. + L_2 e^{(k-3)(\alpha-jT)} e^{-j2a\varphi_0} + \dots e^{-(k-1)ja\varphi_0}\right), \end{aligned} \tag{15}$$

where

$$L_1 = 1 - e^{-(\alpha-jT)+j\varphi_0^{(1)}+ja\varphi_0^{(2)}}; \quad L_2 = 1 - e^{-(\alpha-jT)+j\varphi_0^{(1)}-ja\varphi_0^{(2)}}$$

The sums (15) can be readily evaluated by convolution, and we obtain as final expressions for the coefficients $C_k^{(1)}$ and $C_k^{(2)}$

$$\begin{aligned} C_k^{(1)} &= \frac{jv_0\mu_1}{2(1-a)}\left\{\frac{e^{k(\alpha-jT)} - e^{jk\varphi_0 a}}{e^{\alpha-jT} - e^{j\varphi_0 a}}\left(1 - e^{-(\alpha-jT)+j\varphi_0^{(1)}+ja\varphi_0^{(2)}}\right) + \right. \\ &\quad \left. + e^{-(\alpha-jT)+j\varphi_0^{(1)}+ja\varphi_0^{(2)}+(k-1)ja\varphi_0}\right\} e^{j\omega t_1}, \\ C_k^{(2)} &= \frac{jv_0\mu_1}{2(1+a)}\left\{\frac{e^{k(\alpha-jT)} - e^{-jk\varphi_0 a}}{e^{\alpha-jT} - e^{-j\varphi_0 a}}\left(1 - e^{-(\alpha-jT)+j\varphi_0^{(1)}-ja\varphi_0^{(2)}}\right) + \right. \\ &\quad \left. + e^{-(\alpha-jT)+j\varphi_0^{(1)}-ja\varphi_0^{(2)}-(k-1)ja\varphi_0}\right\} e^{j\omega t_1}. \end{aligned} \tag{16}$$

The resultant coefficients enable us to determine the velocity, the current, and the field intensity of the space charge in any gap of the slow-wave system.

Formulas for the current and power. Using the expressions for the arbitrary constants

(16), let us write down the formula for the convection current in the k-th gap of the system:

$$\widetilde{i}_k = -j\mu_1 I_0 \left\{ \frac{e^{(k-1)(\alpha-jT)+j\varphi^{(1)}}}{1-a^2} - \frac{e^{ja\varphi^{(1)}}}{2a(1-a)} \left[\frac{e^{k(\alpha-jT)} - e^{jk\varphi_0 a}}{e^{\alpha-jT} - e^{j\varphi_0 a}} L_1 + \right.\right.$$
$$\left. + e^{-(\alpha-jT)+j\varphi_0^{(1)}+ja\varphi_0^{(2)}} e^{j(k-1)a\varphi_0} \right] + \frac{e^{-ja\varphi^{(1)}}}{2a(1+a)} \left[\frac{e^{k(\alpha-jT)} - e^{-jk\varphi_0 a}}{e^{\alpha-jT} - e^{-j\varphi_0 a}} L_2 + \right.$$
$$\left.\left. + e^{-(\alpha-jT)+j\varphi_0^{(1)}-ja\varphi_0^{(2)}} e^{-j(k-1)a\varphi_0} \right] \right\} e^{j\omega t_1}. \tag{17}$$

We now calculate the power of interaction between the electron beam and the high frequency field in the k-th gap. For this we make use of the well known form

$$P_k = \frac{1}{2} \frac{v_0}{\omega} \int_0^{\varphi_d} E_k^* \widetilde{i}_k \, d\varphi^{(1)},$$

where

$$E_k^* = \frac{V_1}{d} e^{-\omega t_1} e^{(k-1)(\alpha+jT)} e^{-j\varphi^{(1)}}; \quad \varphi_d \equiv \varphi_0^{(1)}.$$

Substituting into the formula $\widetilde{i}_k$ and E^*_k, we obtain after integration

$$P_k = I_0 V_0 \mu_1^2 \left\{ \frac{-j\varphi_l}{1-a^2} e^{2(k-1)\alpha} - \frac{e^{-j(1-a)\varphi_d} - 1}{2a(1-a)^2} \left[\frac{e^{k(\alpha-jT)} - e^{jk\varphi_0 a}}{e^{\alpha-jT} - e^{j\varphi_0 a}} L_1 e^{(k-1)(\alpha+jT)} + \right.\right.$$
$$\left. + e^{-(\alpha-jT)+j\varphi_d+ja\varphi_0^{(2)}} e^{j(k-1)a\varphi_0} e^{(k-1)(\alpha+jT)} \right] + \frac{e^{-j(1+a)\varphi_d} - 1}{2a(1+a)^2} \left[\frac{e^{k(\alpha-jT)} - e^{-jk\varphi_0 a}}{e^{\alpha-jT} - e^{-j\varphi_0 a}} \times \right.$$
$$\left.\left. \times L_2 e^{(k-1)(\alpha+jT)} + e^{-(\alpha-jT)+j\varphi_d-ja\varphi_0^{(2)}} e^{-j(k-1)a\varphi_0} e^{(k-1)(\alpha+jT)} \right] \right\}. \tag{18}$$

If the system has m gaps, then the total power of interaction for the entire system is determined as the sum of P_k over all the interaction spaces:

$$P_{em} = \sum_{k=1}^{m} P_k.$$

After summation, we obtain the final result

$$P_{em} = I_0 V_0 \mu_1^2 \left\{ \left[\frac{-j\varphi_d}{1-a^2} - \frac{e^{-j(1-a)\varphi_d} - 1}{2a(1-a)^2} \frac{e^{\alpha-jT} - e^{j\varphi_d+ja\varphi_0^{(2)}}}{e^{\alpha-jT} - e^{j\varphi_0 a}} + \right.\right.$$
$$\left. + \frac{e^{-j(1+a)\varphi_d} - 1}{2a(1+a)^2} \frac{e^{\alpha-jT} - e^{j\varphi_d-ja\varphi_0^{(2)}}}{e^{\alpha-jT} - e^{-j\varphi_0 a}} \right] \frac{e^{2m\alpha} - 1}{e^{2\alpha} - 1} + \frac{e^{-j(1-a)\varphi_d} - 1}{2a(1-a)^2} \times$$
$$\times \frac{e^{j\varphi_0 a} - e^{j\varphi_d+j\varphi_0^{(2)}}}{e^{\alpha-jT} - e^{j\varphi_0 a}} \frac{e^{m(\alpha+jT+j\varphi_0 a)} - 1}{e^{\alpha+jT+j\varphi_0 a} - 1} - \frac{e^{-j(1+a)\varphi_d} - 1}{2a(1+a)^2} \times$$
$$\left. \times \frac{e^{-j\varphi_0 a} - e^{j\varphi_d - j\varphi_0^{(2)} a}}{e^{\alpha-jT} - e^{-j\varphi_0 a}} \frac{e^{m(\alpha+jT-j\varphi_0 a)} - 1}{e^{\alpha+jT-j\varphi_0 a} - 1} \right\}. \tag{19}$$

We note that in the case of a self-consistent solution, only the exponential part of the power is meaningful, i.e., the part corresponding to harmonic variation of the current from gap to gap. Let us transform this part of the power to the following form

$$P_{em}^{\exp} = \frac{I_0}{4V_0} V_1^2 \frac{1}{4[\operatorname{ch}(-\theta + j\varphi_0) - \cos\varphi_0 a]} \left\{ \frac{j2\,[\operatorname{ch}(-\theta + j\varphi_0) - \cos\varphi_0 a]}{(1-a^2)\,\varphi_d} \times \right.$$
$$\times \left[\frac{(1+a)\,\overline{M} e^{-j(1-a)\varphi_d} - (1-a)\,\overset{+}{M} e^{-j(1+a)\varphi_d}}{a} - 2 \right] + \frac{\overline{M}^2 e^{j\varphi_0 a} - \overset{+}{M}^2 e^{-j\varphi_0 a}}{a} -$$
$$\left. - \frac{\overline{M}^2 - \overset{+}{M}^2}{2} e^{-j\varphi_0 + \theta} \right\} \frac{e^{2m\alpha} - 1}{e^{2\alpha} - 1}. \tag{20}$$

We use here the notation

$$\frac{\sin(1-a)\,\varphi_{d/2}}{(1-a)\,\varphi_{d/2}} = \overline{M}, \quad \frac{\sin(1+a)\,\varphi_{d/2}}{(1+a)\,\varphi_{d/2}} = \overset{+}{M}.$$

We introduce the additional notation

$$j\left[\frac{(1+a)\,\overline{M} e^{-j(1-a)\varphi_{d/2}} - (1-a)\,\overset{+}{M} e^{-j(1+a)\varphi_{d/2}}}{2a\,(1-a^2)\,\varphi_d} - \frac{1}{(1-a^2)\,d_d}\right] = A_{\varphi_d},$$

$$\frac{\overline{M}^2 e^{j\varphi_0 a} - \overset{+}{M} e^{-j\varphi_0 a}}{a} = M^{(1)}, \quad \frac{\overline{M}^2 - \overset{+}{M}^2}{a} = M^{(2)}.$$

Let us explain the physical meaning of the coefficients introduced. The coefficient $A_{\varphi d}$ can be represented in the form

$$A_{\varphi_d} = \left\{ \frac{2}{(1-a^2)^2} - \frac{\cos(1-a)\,\varphi_d}{2a\,(1-a)^2} + \frac{\cos(1+a)\,\varphi_d}{2a\,(1+a)^2} + \right.$$
$$\left. + j\left[\frac{\sin(1-a)\,\varphi_d}{2a\,(1-a)^2} - \frac{\sin(1+a)\,\varphi_d}{2a\,(1+a)^2} - \frac{\varphi_d}{1-a^2}\right]\right\} = \frac{4V_0}{I_0 V_1^2} P_{e\text{д}}.$$

Thus, in the power formula the coefficient $A_{\varphi d}$ describes the diode effect [12] due to the finite transit angle of the electrons through the gap. When $2\pi < \varphi_d < 3\pi P_{di} < 0$, i.e., when $2\pi < \omega_n d/\nu_0 < 3\pi$ where ω_n is the natural frequency of each cell of the system, monotron type oscillations can occur in the tube. The coefficients $M^{(1)}$ and $M^{(2)}$ can be rewritten in the form

$$M^{(1)} = j\,(\overline{M}^2 + \overset{+}{M}^2)\frac{\sin\varphi_0 a}{a} + \frac{16\,V_0}{I_0 V_1^2} P_{ead} \cos\varphi_0 a,$$

$$M^{(2)} = \frac{16\,V_0}{I_0 V_1^2} P_{ead}$$

These are the coefficients of beam modulation in the gap, with account of the diode effect.

Using the notation introduced, we write

$$P_{em}^{\exp} = \frac{I_0}{4V_0} V_1^2 \frac{1}{4[\operatorname{ch}(-\theta + j\varphi_0) - \cos\varphi_0 a]} \times$$
$$\times \{4\,[\operatorname{ch}(-\theta + j\varphi_0) - \cos\varphi_0 a]\,A_{\varphi_d} + M^{(1)} - M^{(2)} e^{\theta - j\varphi_0}\} \frac{e^{2m\alpha} - 1}{e^{2\alpha} - 1}. \tag{21}$$

2. LINE EQUATION WITH ACCOUNT OF THE INDUCED CURRENT

The most rigorous method of analysis of transmission lines and all other microwave circuits, with account taken of their finite dimensions, is the method based on solving the Maxwell equations with boundary conditions imposed by the construction of the system. This method can be used, however, only when the boundary conditions in the system are sufficiently simple (more accurately, when the boundary is described with sufficient simplicity in terms of the coordinates).

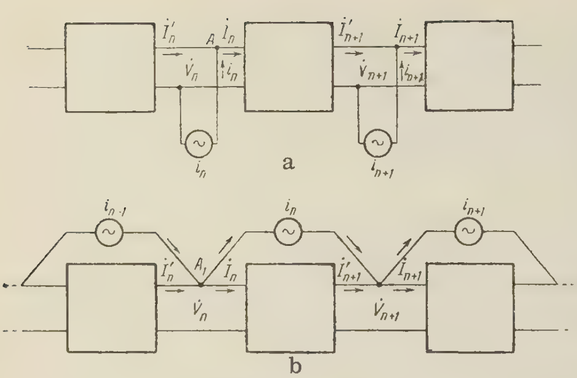


Fig. 2.

An investigation of microwave systems with the aid of Maxwell's equations shows that there can exist in such systems an infinite number of oscillation modes (TM, TE, TEM). In the slow-wave systems of interest to us, however, types TM_{01} (E_{01}) modes are used. The higher E modes have a high frequency and, therefore, do not propagate within the frequency band under consideration. H modes are not excited by the electron beam. All this enables us to use, in the calculation of certain periodic systems, equivalent circuits with lumped constants, so that the main shortcoming of this method of calculation, namely, that it describes the behavior of the system at only one oscillation mode, becomes immaterial. The equivalent-circuit elements can be calculated from the geometrical dimensions of the slow-wave systems or can be determined from data obtained by "cold" measurements. In some cases it is convenient to use equivalent circuits made up of segments of long lines [13].

Line equation. In the derivation of the line equation we recognize that there are two ways of setting up the equivalent circuit of the line with respect to the electron beam. Both ways are shown in Figs. 2a, b. Let us derive the line equation for each version, applying the two-port theory developed in [14].

As is well known from two-port theory, the input and output voltages and currents are related as follows:

$$\begin{pmatrix} \dot V_n \\ \dot I_n \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \dot V_{n+1} \\ \dot I'_{n+1} \end{pmatrix}, \tag{22}$$

where

$$A_{11} = \operatorname{ch}\theta_1;\ A_{12} = Z_0 \operatorname{sh}\theta_1;\ A_{21} = \frac{\operatorname{sh}\theta_1}{Z_0};\ A_{22} = \operatorname{ch}\theta_1;$$

θ_1 is the transfer functions of the two-ports; Z_0 is the characteristic impedance of the symmetrical two-port.

We transform (22) as follows:

$$\begin{pmatrix} \dot I_n \\ \dot I'_{n+1} \end{pmatrix} = \frac{1}{Z_0} \begin{pmatrix} \operatorname{cth}\theta_1 & -\dfrac{1}{\operatorname{sh}\theta_1} \\ \dfrac{1}{\operatorname{sh}\theta_1} & -\operatorname{cth}\theta_1 \end{pmatrix} \begin{pmatrix} \dot V_n \\ \dot V_{n+1} \end{pmatrix}. \tag{23}$$

Using (23), we establish the connection between the induced current and the circuit voltage for the case shown in Fig. 2a. We assume here that the quantities vary as follows:

$$i_{n+1} = i_n e^{-\theta},\ \dot V_{n+1} = \dot V_n e^{-\theta},\ \dot I'_{n+1} = \dot I_n e^{-\theta}.$$

We employ Kirchhoff's law for currents at the point A.

$$i_n = \dot I_n - \dot I'_n = \dot I_n - \dot I_{n+1} e^{\theta}. \tag{24}$$

From (24) and (23) we get

$$i_n = \frac{2(\operatorname{ch}\theta_1 - \operatorname{ch}\theta)}{Z_0 \operatorname{sh}\theta_1} \dot V_n. \tag{25}$$

Using (25) we determine the power released in the circuit of the k-th gap:

$$P_k = \frac{\dot{V}_k^* i_k}{2} = \frac{2(\mathrm{ch}\,\theta_1 - \mathrm{ch}\,\theta)}{Z_0\,\mathrm{sh}\,\theta_1} V_1^2 e^{2(k-1)\alpha}$$

The total power released in m gaps is determined by summing P_k over all the gaps

$$P_{lm} = \frac{2(\mathrm{ch}\,\theta_1 - \mathrm{ch}\,\theta)}{Z_0\,\mathrm{sh}\,\theta_1} V_1^2 \frac{e^{2m\alpha} - 1}{e^{2\alpha} - 1}. \tag{26}$$

For the case shown in Fig. 2b, using Kirchhoff's law for currents at the point A_1 we get

$$i_n - i_{n-1} = i_n(1 - e^{\theta}) = I_n' - I_n. \tag{27}$$

The expressions for the induced current and for the power released in the line by the induced current have for this case the form

$$\dot{i}_n = -\frac{2(\mathrm{ch}\,\theta_1 - \mathrm{ch}\,\theta)}{Z_{01}\,\mathrm{sh}\,\theta_1 (1 - e^{-\theta})}, \tag{28}$$

$$P_{lm} = \frac{(2\,\mathrm{ch}\,\theta_1 - \mathrm{ch}\,\theta)}{Z_{01}\,\mathrm{sh}\,\theta_1 (1 - e^{-\theta^*})(1 - e^{\theta})} V_1^2 \frac{e^{2m\alpha} - 1}{e^{2\alpha} - 1}. \tag{29}$$

In order to avoid writing two different dispersion equations for cases a and b (Fig. 2), we introduce the generalized characteristic impedance Z_0^e, which for case a has a value Z_0, and for case b

$$Z_0^e = Z_0(1 - e^{-\theta^*})(1 - e^{\theta}).$$

Using (22), we write down the boundary conditions for the waves in the line; we have in mind here the circuit of Fig. 3.

The input conditions are

$$\frac{\dot{I}_0 Z_{02} + \dot{V}_0}{\dot{V}_1} = \mathrm{ch}\,\theta_i + \mathrm{sh}\,\theta_i. \tag{30}$$

The output conditions are

$$\frac{V_m}{V_{m+1}} = \mathrm{ch}\,\theta_i + \frac{Z_0}{Z_{m+1}}\,\mathrm{sh}\,\theta_i. \tag{31}$$

Let us dwell now on the concept of the coupling resistance R_c. The coupling resistance is defined as

$$R_c = \frac{E_1^2}{2\beta^2 P} = \frac{V_1^2}{2P},$$

where P is the power flowing through the system.

It is easy to show (for example for equivalent circuits in the form of a chain of pi-connected two-ports) that

$$R_c = \frac{Z_0\,\mathrm{sh}\,\theta_1}{(e^{-\theta} - 1)(e^{-\theta^*} - 1)}.$$

We shall use the last expression to compare the cascade analysis with the harmonic analysis.

3. DISPERSION EQUATION AND BOUNDARY CONDITIONS

In the preceding sections we gave relations for the power of interaction between the electrons and the field of the line, and also the power released in the line as a result of the induced current. The condition for the compatibility of the resulting solutions can obviously be represented in the form

$$P_{em}^{\mathrm{exp}} = -P_{lm}. \tag{32}$$

which should be in identity if we insert in the expressions for $P_{em}{}^{\mathrm{exp}}$ and P_m all the values of the parameters characterizing our circuit, namely φ_0, θ_1, φ_d, I_0, V_0, $Z_0{}^e$, a, and θ. The only unknown is θ. We can therefore regard Eq. (32) as an equation for the determination of θ.

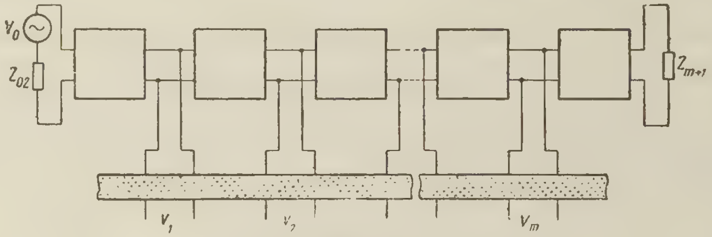


Fig. 3

The roots of this equation, which make (32) an identity, will obviously be the transfer constants characterizing our system. Consequently the equation resulting from (32) can be called the dispersion equation for the given TWT circuit.

Using (21) and (26) we obtain the dispersion equation

$$\left(\mathrm{ch}\,\theta - \mathrm{ch}\,\theta_1 - \frac{I_0 Z_0}{8V_0} A_{\varphi_d}\right)[\mathrm{ch}\,(-\theta + j\varphi_0) - \cos a\varphi_0] = \\ = \frac{1}{8}\,\frac{I_0 Z_0^e}{4V_0}\,\mathrm{sh}\theta_1\,(M^{(1)} - M^{(2)} e^{\theta - j\varphi_0}). \tag{33}$$

In the case of infinitesimally narrow gaps ($\varphi_d \rightarrow 0$), we obtain the following equation

$$(\mathrm{ch}\,\theta - \mathrm{ch}\,\theta_1)\,[\mathrm{ch}\,(-\theta + j\varphi_0) - \cos a\varphi_0] = \\ = \frac{1}{2}\,\frac{I_0 Z_0^e}{4V_0}\,\mathrm{sh}\,\theta_1 j\varphi_0\,\frac{\sin a\,\varphi_0}{a\varphi_0}. \tag{34}$$

Eq. (34) is transformed in the case of a slow-wave system of the coupled resonator type, into Eqs. (46) and (47) of [6], where the operation of a TWT with such a slow-wave system is analyzed.

Eq. (33) is a fourth-degree equation with respect to the transfer coefficient $K = e^{-\theta}$:

$$AK^4 + BK^3 + CK^2 + DK + E = 0;$$

thus, the field in the system is described by a superposition of four partial waves, corresponding to four transfer coefficients, obtained by solving the equation

$$V_m = \sum_i^4 V_{1i} e^{-(m-1)\theta_i}.$$

Using the properties of the roots of the equation, we can determine the character of these waves:

$$e^{-\theta_1} e^{-\theta_2} e^{-\theta_3} e^{-\theta_4} = \frac{E}{A},$$

$$E = \frac{1}{4}\{1 + C^3\,\mathrm{sh}\,\theta_1\,[4A_{\varphi_d} - M^{(2)}]\}\,e^{-j\varphi_0},$$

$$A = \frac{1}{4}[1 + 4A_{\varphi_d} C^3\,\mathrm{sh}\,\theta_1]\,e^{j\varphi_0}, \qquad \frac{E}{A} = \dot{K} e^{-j2\varphi_0}.$$

In the case when $\varphi_d \rightarrow 0$ we have

$$\dot{K} = 1, \quad \theta_1 + \theta_2 + \theta_3 + \theta_4 = 2j\varphi_0. \tag{35}$$

If follows from (35) that one of the four waves has a negative propagation constant (backward wave). It also follows from (35) that if there are growing waves in the system, there exist also symmetrically damped waves, since the sum of the growth exponents is equal to zero.

For finite values of φ_d we have

$$\theta_1 + \theta_2 + \theta_3 + \theta_4 = 2j\varphi_0 + \dot{r},$$

i.e., the symmetrical wave is disturbed by the diode effect.

Let us write expressions for $\tilde{v}$ and $\tilde{i}$, for each wave component with $m = 1$ and $\varphi = 0$. We put here $-\theta_i = j\varphi_0 \div \Delta_i$:

$$\tilde{v}_i^{\exp} = jv_0 \frac{V_{1i}}{4V_0} \frac{(1 - \cos\varphi_d) - j\sin\varphi_d}{\varphi_d(1 - a)} \frac{\Delta_i}{\Delta_i^2 + (\varphi_0 a)^2}, \tag{36}$$

$$\tilde{i}_i^{\exp} = -I_0 \frac{V_{1i}}{4V_0} \frac{\varphi_0[(1 - \cos\varphi_d) - j\sin\varphi_d]}{\varphi_d(1 - a^2)} \frac{1}{\Delta_i^2 + (\varphi_0 a)^2}. \tag{37}$$

Expressions (36) and (37) are obtained under the assumption $|\Delta_i| \ll 1$ and $a \ll 1$, i.e., under the assumption that the deviation from synchronism is small and that the current densities are not too large; when φ_d is small, the second limitation drops out.

The boundary conditions (10), written in terms of the wave components, have the form

$$\sum_i^4 \frac{V_{1i}\Delta_i}{\Delta_i^2 + (a\varphi_0)^2} = 0, \qquad \sum_i^4 \frac{V_{1i}}{\Delta_i^2 + (a\varphi_0)^2} = 0. \tag{38}$$

Supplementing these boundary conditions by conditions (30) and (31), we have

$$\sum_i^4 V_{1i}e^{\theta_i} = \tilde{I}_0 Z_{02} + \tilde{V}_0, \qquad \sum_i^4 V_{1i}e^{-m\theta_i} \operatorname{sh}\theta_1 \left(1 - \frac{Z_0}{Z_{m+1}}\right) = 0. \tag{39}$$

We obtained a system of boundary conditions (38) and (39), which enable us to determine the initial amplitudes of the four partial waves. Knowledge of these amplitudes makes it possible to calculate the distribution of the current and voltage along the tube and the gain produced by the latter.

4. COMPARISON WITH RESULTS OF HARMONIC ANALYSIS AND LIMITS OF APPLICABILITY OF ORDINARY HARMONIC ANALYSIS

By ordinary harmonic analysis we mean here an analysis in which the field in the line is determined from (11), i.e., we consider the interaction between the electrons and the field of only one spatial harmonic. This analysis, obviously, is perfectly valid for continuous interaction. Let us consider Eq. (34) for the case of continuous interaction.

We put

$$\left\{\begin{array}{l} \theta = \Gamma l, \\ \theta_1 = \Gamma_1 l, \\ \varphi_0 = \beta_e l \end{array}\right\}$$

and assume that θ and $\theta_1 \to 0$ because $l \to 0$.

Dividing both halves of (34) by l^4 and going to the limit as $l \to 0$, we obtain

$$(-\Gamma + j\beta_e)^2 \cdot \frac{2jC^3\Gamma^2\Gamma_1\beta_e}{(\Gamma_1^2 - \Gamma_\beta^2)} - \beta_p^2.$$

Introducing the notation $4QC^3 = (\beta_p/j\Gamma)^2$, we obtain the Pierce equation [1] for finite C:

$$(-\Gamma + j\beta_e)^2 = \frac{2jC^3\Gamma^2\Gamma_1\beta_e}{\Gamma_1^2 - \Gamma^2} + 4Q\Gamma^2C^3.$$

Thus, for the case of continuous interaction, both analyses yield identical results.

Let us consider now the case of discrete interaction, i.e., θ_1, θ, $\varphi_0 \neq 0$. For simplicity we consider the case of small C, when transition from the equation in $e^{-\theta}$ can be changed to equations with respect to $\Delta = j\varphi_0 - \theta$ by using the expansion $e^{-\theta} = e^{-j\varphi_0}(1 + \Delta)$, which is valid for small Δ (or small C, since Δ is of the same order as C).

We put

$$-\theta = -j\varphi_0 + \Delta, \qquad -\theta = -\theta_1 + \Delta + \vartheta, \qquad -\theta_1 = j\varphi_0 - \vartheta.$$

Here ϑ is also of the same order as C.

From Eq. (33) we get

$$\begin{aligned}
&\Delta^4\left(-jA_{\varphi_d}C^3\cos^2\frac{\varphi_0}{2} - \frac{1}{2}\operatorname{cth}\theta_1\right) + \\
&+\Delta^3\left(\vartheta\operatorname{cth}\theta_1 - 1 - jA_{\varphi_d}4C^3\sin\varphi_0 + M^{(2)}2C^3\cos^2\frac{\varphi_0}{2}\right) + \\
&+\Delta^2\left(\operatorname{cth}\theta_1\, 2\sin^2 a\frac{\varphi_0}{2} + \frac{\vartheta^2}{2}\operatorname{cth}\theta_1 - \vartheta - 8\sin^2\frac{\varphi_0}{2}A_{\varphi_d}C^3 - \right.\\
&\left. - j4C^3\sin\varphi_0 M^{(2)} - jC^3A_{\varphi_d}2\sin^2 a\varphi_0 + M^{(1)} - M^{(2)}\right) + \\
&+\Delta\left[4\vartheta\sin^2 a\frac{\varphi_0}{2}\operatorname{cth}\theta_1 - 4\sin^2 a\frac{\varphi_0}{2} - C^3M^{(2)}\sin^2\frac{\varphi_0}{2} - \right.\\
&\left. - j8C^3\left(2\sin^2 a\frac{\varphi_0}{2}A_{\varphi_d} + M^{(1)} - M^{(2)}\right)\right] + \left(\vartheta^2 2\sin^2 a\frac{\varphi_0}{2}\operatorname{cth}\theta_1 - \right.\\
&\left. - 2\vartheta 2\sin^2 a\frac{\varphi_0}{2} - 8C^3A_{\varphi_d}\sin^2\frac{\varphi_0}{2}\sin^2 a\frac{\varphi_0}{2} + M^{(1)} - M^{(2)}\right) = 0.
\end{aligned} \tag{40}$$

Subject to the condition $C\,|\operatorname{cth}\theta_1| \ll 1$, we can rewrite (40) as

$$(\delta + jb + d)(\delta^2 + 4QC) + \left(\frac{\sin\frac{\varphi}{2}}{\varphi/2}\right)^2 \frac{\sin\varphi_0 a}{\varphi_0 a}\,\frac{M^{(1)} - M^{(2)}}{2\varphi_0} = 0, \tag{41}$$

where

$$\delta = \frac{\Delta}{\varphi_0 C}; \quad 4QC = \frac{a^2}{C^2}; \quad b = \operatorname{Im}\frac{\vartheta}{\varphi_0 C}; \quad d = \operatorname{Re}\frac{\vartheta}{\varphi_0 C}$$

are the parameters used in [1];

$$\frac{M^{(1)} - M^{(2)}}{2\varphi_0} = \frac{\sin^2 a\,\frac{\varphi_0}{2}\,[\overline{M}^2 - \overset{+}{M}{}^2]}{\varphi_0 a} + j\frac{\sin\varphi_0 a}{\varphi_0 a}\,|M^2 + \overset{+}{M}{}^2|.$$

Taking into account the smallness of a, Eq. (41) assumes the form

$$(\delta_1 + jb_1 + d_1)(\delta_1^2 + 4Q_1C_1) + j = 0. \tag{42}$$

Eq. (42) is the Pierce equation for small C. The quantities C_1, δ_1, d_1, b_1, and Q_1 include an allowance for the amplitude factor of the spatial harmonic $\dfrac{\sin\frac{\varphi_0}{2}}{\varphi_0/2}$. Thus, within the framework of the condition

$$C\,|\operatorname{cth}\theta_1| \ll 1 \tag{43}$$

the cascade analysis does not differ from the harmonic analysis. On the other hand, at the limit of the pass band and beyond its limits, these analyses are essentially different. For comparison, let us consider a tube operating at frequencies corresponding to the upper cutoff frequency or lying above this frequency, for systems with normal dispersion, or at frequencies respectively equal to the lower cutoff frequency and lying below it, for systems with anomalous dispersion. For simplicity we consider the case of zero losses and infinitesimally narrow gaps. At the frequencies chosen, the phase shift of the field per cell is $\operatorname{Re}\theta \simeq \pi$. Let us assume that the condition of synchronism between the electrons and the fields is satisfied, i.e., $\varphi_0 \simeq \pi$. In this case the synchronism condition is simultaneously satisfied for two spatial harmonics of the field — the forward and backward neighboring wave numbers. With this, the harmonic analysis leads to the necessity for simultaneous solution of two fourth-degree equations (corresponding to the forward and backward harmonics), while the coupling resistance is infinite in the case of zero losses.

The cascade analysis yields for this case the following simple equation, which follows from (34) for the case under consideration:

$$(\delta^2 - \Delta_2^2)\,[(\delta - \Delta_1)^2 - q^2] + 1 = 0, \tag{44}$$

where

$$\delta = \frac{-j\,\theta - \pi}{A}; \quad \Delta_1 = \frac{\varphi_0 - \pi}{A}; \quad \Delta_2 = \frac{-j\,\theta_1 - \pi}{A}; \quad q = \frac{a\,\varphi_0}{A};$$

$$A^4 = \frac{I_0 Z_0^2}{2V_0}\varphi_0\,\frac{\operatorname{sh}\theta_1}{j}\,\frac{\mathrm{sih}\,a\,\varphi_0}{a\,\varphi_0}.$$

The foregoing difference is due to the fact that the mechanism of amplification of the TWT is different within the passband and outside it. The amplification is not connected here with the power flux of the high frequency field in the system, as takes place in the center of the band; in the region under consideration, the system properties approach those of a set of uncoupled resonating elements, where energy is transferred only by the electron beam. Thus, the mechanism of amplification is more similar to the klystron nature than to the customary one for the TWT. On the other hand, the cascade analysis describes equally well both the klystron mode and the traveling-wave mode, thereby making it preferable for the description of a tube of the filter type, in which both operating modes exist.

CONCLUSION

The dispersion equation and the boundary conditions given in this article enable us to calculate the frequency characteristics of tubes of the filter type in the amplification mode or in the generation mode (BWT). It becomes possible here to examine the characteristics of the TWT at the limit of the passband of the system, and beyond its limits, and to study in detail the features of TWT amplification in these regions.

We note that the one-dimensional approximation used in the solution of our problem does not lead to serious limitations on the theory, since the results can be readily generalized to the case of thick beams by introducing a plasma-frequency reduction coefficient

$p = 1\Big/\sqrt{1 + \frac{T_1^2}{\gamma_0^2}}$ [15], which takes into account the effect of the conducting walls in the drift space on the bunching process in the beam. In this case the smaller the lengths of the discontinuities in the drift spaces at the points of interaction compared with the plasma wavelength the less they influence the correction made.

The transverse distribution of the field in the interaction spaces, in the case of gridless gaps, can be taken into account by using the known averaging coefficients for sufficiently narrow gaps, namely

$$\frac{2I_1(r_1')}{r_1'\,I_0(r_2')}$$

for cylindrical gaps and

$$\frac{\operatorname{sh} C}{C \operatorname{sh} A}$$

for slit gaps. Here r'_1 — radius of the beam; r'_2 — radius of the drift tube; C — half the beam width; A — half the slit width (C and A are in radians).

REFERENCES

1. J.R. Pierce. Traveling Wave Tubes. Translated from the English. Soviet Radio Press, 1950.
2. L.A. Vaynshteyn. Electron waves in periodic structures. ZhTF 1957, 27, 10, 2340.
3. L.N. Loshakov. Application of the Lorentz lemma for the determination of the propagation constants in the interaction between an electron beam and spatial harmonics. Radiotekhnika i elektronika, 1957, 2, 4, 461.
4. M.F. Stel'makh. On the interaction of an electron beam with the field of spatial harmonics. ibid. 1957, 2, 4, 470.
5. E.J. Nalos. A hybrid type traveling-wave tube for high-power pulsed amplification. IRE Trans. Electron Devices, 1958, 5, 3, 161.
6. R.W. Gould. Characteristics of traveling-wave tubes with periodic circuits. IRE Trans. Electron Devices, 1958, 5, 3, 186.
7. V.N. Shevchik, Yu.D. Zharkov. Cascade bunching of electrons as applied to the analysis of the interaction between an electron beam and a traveling electromagnetic wave, Radiotekhnika i elektronika, 1957, 2, 2, 237.
8. V.N. Shevchik, L.D. Pokrovskiy. Energy exchange in discrete interaction between electrons and a traveling electromagnetic wave. Izvestiya Vuzov, Ministry of Higher Education USSR — Radiophysics, 1960, 3, 3, 155.
9. Hsieh Chia-lin. Wave space-charge of multiple bunching. Scientia Sinica, 1957, 6, 1, 33.
10. J.R. Pierce, N. Wax. A note on filter type traveling-wave amplifiers, Proc. IRE, 1949, 37, 6, 622.
11. V.N. Shevchik, L.Ya. Mayofis. Characteristic equations of traveling wave tubes and backward wave tubes in discrete interaction between electrons and electromagnetic waves. Izv. Vuzov MVO SSSR, Radiotekhnika, 1959, 2, 3, 367.
12. V.N. Shevchik. Principles of microwave electronics, Soviet Radio Press, 1959.
13. Hirano Junzo. Characteristics of interdigital circuits and their use for amplifiers, Proc. Instn. Electr. Engrs., 1958, B105, Suppl. No. 11, 780.
14. N.N. Krylov. Theoretical Foundations of Radio Engineering. Marine Transport Publishing House, 1953.
15. A.L. Cullen, I.M. Stephenson. An experimental investigation of velocity-modulated electron beams. Proc. Instn. Electr. Engrs., 1958, B105, Suppl. No. 12, 843.
16. D.G. Dow. Behavior of TWT near circuit cut off. IRE Trans. Electron Devices, 1960, 7, 3, 123.

Received by editor 10 December 1960

PROPAGATION OF MICROWAVE OSCILLATIONS IN A WAVEGUIDE CONTAINING THE CATHODE PARTS OF A GLOW DISCHARGE

A.M. Starik

The dependence of the phase shift and attenuation produced by the cathode parts of a glow discharge in a rectangular waveguide on the pressure of helium and neon are investigated. The effect of transverse dimensions of the waveguide on the course of these dependences is also determined. The data obtained are interpreted by starting from the presence of an inverse proportionality between the length of the cathode region and the pressure.

INTRODUCTION

Many problems in waveguide techniques necessitate the use of electrically controlled modulators, attenuators, switches, and other devices intended to change the amplitude, phase, or direction of propagation of centimeter waves in waveguides. These problems can be solved by using ferrites or gas-discharge plasma. Devices of this kind, using gas-discharge plasma which fills a more or less extensive portion of the waveguide, have been treated in several papers [1 - 8]. The theory developed in some of these papers pertains to the case when the waveguide is uniformly filled with electron gas, and can yield only very general recommendations concerning devices in which the cathode parts of a glow discharge are used. The published experimental results give no idea of the dependence of the characteristics of such devices on their construction and the substance used in them.

We give below some data obtained with the aid of models, in which a cold cathode of considerable length was situated inside a waveguide.

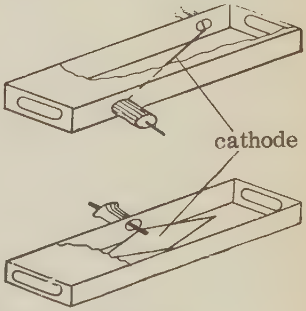


Fig. 1. Construction of experimental models

1. EXPERIMENTAL CONDITIONS

The experimental models were made in the form of waveguide segments, to the ends of which were sealed resonant glass windows of very low Q. Inside the waveguide, in a plane parallel to its broad wall, was installed a cathode of one of two constructions (Fig. 1). In the first case the cathode was a round wire (0.4 mm in diameter) stretched at an angle to the waveguide axis, so as to ensure matching. A cathode of this type was proposed by Pringle for use in a gas-discharge phase shifter [2]. The second cathode construction comprised a diamond-shaped plate (0.4 mm).

During the course of the investigation it proved necessary to carry out experiments with devices made of waveguides with nonstandard cross section. Quarter-wave transformers were mounted in these devices near the windows.

The characteristics of the devices described depend appreciably on the quality of their outgassing during evacuation. In this connection, the evacuation was carried out for many hours at about 400°C, after which a getter was sputtered in the equipment, and conditioning was repeated many times with subsequent refilling. After the final pressure was established in the device, additional conditioning was carried out.

The measuring apparatus (Fig. 2) was used to determine the dependence of the phase shift and of the attenuation produced by the device on the pressure of neon or helium. The phase

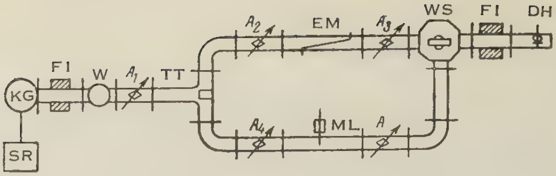


Fig. 2. Measuring scheme.

SR — stabilized rectifier; A — attenuator; KG — klystron generator; ML — measuring line; FI — ferrite insulator; W — wavemeter; EM — experimental model; WS — waveguide switch; TT — twin tee junction; DH — detector head

shift was determined from the change in the position of the standing-wave node, observed with the aid of a slotted line [5], while the attenuation was measured by the substitution method.

A check on the reproducibility of the measurement results made with different models at different times under identical conditions has shown that the spread in the phase-shift values did not exceed 10%. The course of the variation of attenuation with pressure was reproduced with sufficient accuracy, but the absolute values of the attenuation were subject to somewhat greater error than the phase-shift values. This was found to be due to insufficient outgassing of the devices, which affected the attenuation more than the phase shift.

2. RESULTS OF MEASUREMENTS ON MODELS WITH FLAT CATHODE

Figs. 3 and 4 show the dependence of the attenuation L and the phase shift θ on the pressure of neon. The total cathode area, with maximum width 10 mm and length about 80 mm, was 8.65 cm^2. The data given pertain, therefore, to a cathode current density j_k = 1.4 ma/cm^2; the standing-wave coefficient in these models, made of a waveguide with 10 × 23 mm cross section, in a range $\geq 10\%$, did not exceed 1.15 in the absence of a discharge. As can be seen from the curves, the dependence of the attenuation and of the phase shift curves exhibit a sharp maximum at one and the same pressure, 3 mm Hg.

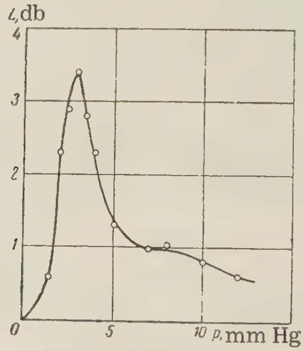


Figure 3

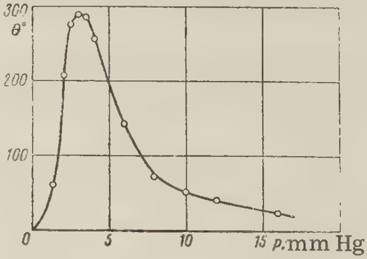


Figure 4

Fig. 3. Dependence of attenuation on the pressure of neon (flat cathode), I = 12 ma.

Fig. 4. Dependence of phase shift on the neon pressure (flat cathode), I = 12 ma.

Analogous variations of L and θ with the pressure of neon were obtained with models made of 4 × 23 mm waveguides. To compare the results obtained with waveguides having different cross sections, we present Table 1, which indicates the pressure values corresponding to the maxima of the curves (p_{max}), and also the values of the phase shift at the maximum in both cases. The interelectrode distance h given in the table is determined as the

Table 1

(Flat cathode, I = 12 ma, j_k = 1.4 ma/cm^2, neon)

Waveguide dimensions, mm		Interelectrode distance h, mm	p_{max}, mm Hg		$\theta°_{max}$	h_1/h_2	p_{max_2}/p_{max_1}		$\theta_{max_1}/\theta_{max_2}$
a	b		Based on L	Based on θ			Based on L	Based on θ	
23	10	4.8	3.0	3.0	290	2.7	3.0	1.8	1.0
23	4	1.8	9.0	5.5	300				

distance between the internal surface of the broad wall of the waveguide to the nearest cathode surface: h = (b - t)/2 (t is the thickness of the cathode).

The data of Table 1 show that p_{max}, determined from the attenuation curve, varies in inverse proportion to the interelectrode distance; p_{max}, determined from the phase-shift curve, has a weaker variation. θ_{max} remains constant.

3. RESULTS OF MEASUREMENTS IN MODELS WITH CYLINDRICAL CATHODE

a) Neon. Figs. 5, 6, and 7 show the dependence of the phase shift, attenuation, and discharge voltage drop U_d on the pressure at various discharge currents. Fig. 8 illustrates the pressure dependence of the quantity F = 100 L/θ, where L is in decibels and θ is in degrees.

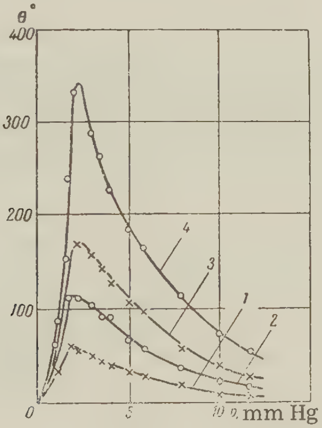


Figure 5

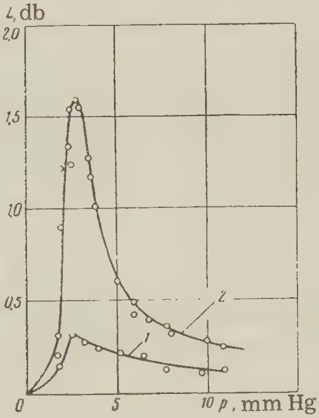


Figure 6

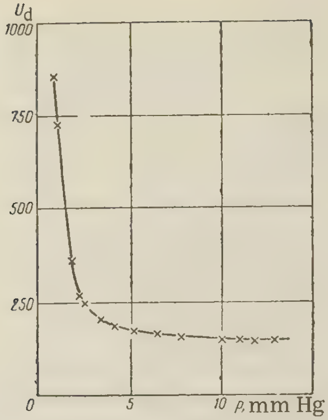
Fig. 5. Dependence of the phase shift on the neon pressure (cylindrical cathode): 1 - I = 2 ma; 2 - I = 3 ma; 3 - I = 4 ma; 4 - I = 5.5 ma

Fig. 6. Dependence of attenuation on the neon pressure (cylindrical cathode: 1 - I = 4 ma; 2 - I = 5.5 ma

According to Figs. 5 and 6, a sharp maximum occurs on the curves of L and θ, as in the case of a flat cathode, but shifted somewhat towards lower pressures. The function E(p) differs in having a minimum at 5.5 — 6 mm Hg.

The results obtained with models made of 4 × 23 mm waveguides are compared in Table 2 with the earlier results. The attenuation maximum shifted here more strongly than in the case of a flat cathode; on the other hand the shift in the phase-shift maximum decreased somewhat.

b) Helium. Fig. 9 shows the dependence of F on the pressure of helium. The pressure variations of L, θ, and U_d are similar in character to the analogous variations obtained for neon.



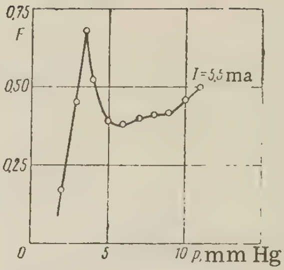


Fig. 7. Dependence of the voltage drop across the discharge on the neon pressure (cylindrical cathode)

Fig. 8. Dependence of parameter F on the neon pressure (cylindrical cathode)

However, whereas in neon the phase shift at the maximum, for a current of 4 milliamperes, is 1.6 times smaller than for helium, the ratio of the maximum attenuations is $L_{max}^{He}/L_{max}^{Ne} = 12 - 13$. The pressure corresponding to the maximum attenuation for helium at $I = 4$ milliamperes is 6.5 mm Hg. Measurements have shown that when $I > 4$ ma, a sharp increase in the SWR sets in at the maximum of the attenuation curve.

The phase-shift curve has a maximum at $p = 5.5$ mm Hg. Unlike in neon, the dependence of F on p has no minimum, and the values of F are much higher.

A comparison with the results obtained in models made of 4 × 23 mm waveguides is shown in Table 3. As in the preceding cases, the position of the attenuation maximum

Table 2

(Cylindrical cathode, $I = 4$ ma, $j_k = 1.4$ ma/cm^2, neon)

Waveguide dimensions, mm		Interelectrode distance h, mm	p_{max}, mm Hg		$\theta°_{max}$	h_1/h_2	p_{max_2}/p_{max_1}		$\theta_{max2}/\theta_{max_1}$
a	b		Based on L	Based on θ			Based on L	Based on θ	
23	10	4.8	2.7	2.3	170	2.7	4.4	1.5	1.1
23	4	1.8	12	3.5	180				

Table 3

(Cylindrical cathode, $I = 4$ ma, $j_k = 4$ ma/cm^2, helium)

Waveguide dimensions, mm		Interelectrode distance h, mm	p_{max}, mm Hg		$\theta°_{max}$	h_1/h_2	p_{max_2}/p_{max_1}		$\theta_{max_2}/\theta_{max_1}$
a	b		Based on L	Based on θ			Based on L	Based on θ	
23	10	4.8	6.5	5.5	270	2.7	2.3	1.0	0.5
23	4	1.8	15.0	7.0	130				

changes here noticeably, and p_{max} varies approximately in inverse proportion to the interelectrode distance. The maximum of the phase shift is displaced slightly, and its value is reduced by one-half.

4. DISCUSSION OF RESULTS

It is known [9] that the distribution of the concentration of electrons n over the length of the discharge gap is characterized by a curve with sharp maximum in the region of negative glow discharge. According to the data of [10], in hydrogen, for example, the concentration at the maximum is 300 times greater than in the positive column. In the case when the waveguide is filled with the positive column of a glow discharge, the phase shift and the attenuation coefficient should increase monotonically with increasing pressure in the range of pressures where $\nu_{col} < \omega$ (ν_{col} is the frequency of collision of electrons with the atoms of the gas, and ω is the circular frequency) [7]. The sharp maxima on the curves L(p) and θ(p) can therefore be ascribed to the presence of the cathode region in the waveguide. Indeed, as the pressure increases, ν_{col} increases, and consequently the concentration of electrons necessary to maintain the specified current also increases. At the same time, however, the length of the cathode region decreases in accordance with the following law [9] (for the plane case)

$$pd = B = \text{const.} \qquad (1)$$

This shrinking of the cathode parts is schematically illustrated by Fig. 10, where the dashed line delineates the boundary of the cathode parts. The smaller the portion of the waveguide cross section occupied by the region with a high electron concentration, the weaker the interaction with the microwave signal. It is obvious that there should exist an optimum position of the maximum of high concentration relative to the cathode, corresponding to the maximum interaction with the microwave signal. At high pressures the maximum of n approaches the cathode, and the interaction weakens. This circumstance can indeed explain the right-hand branch of the curves L(p) and θ(p). The left-hand branch, as in the case of the positive column, is due to the increased electron concentration with increasing pressure.

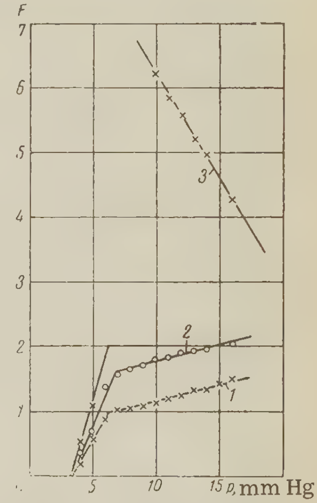


Fig. 9. Dependence of the parameter F on the helium pressure (cylindrical cathode)
1 — I = 3 ma; 2 — I = 4 ma; 3 — I = 5.5 ma.

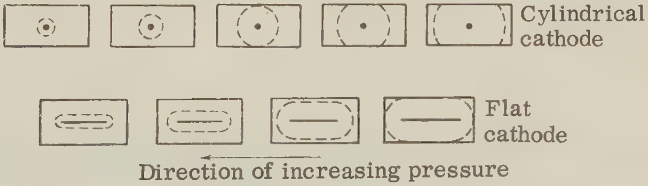


Fig. 10. Change in extent of the cathode region

The optimal position of the maximum of high concentration is reached at a pressure for which the length of the cathode region d amounts to a definite portion η of the interelectrode distance, which is a fraction $\varkappa$ of the height of the waveguide b, i.e.,

$$d = \eta\varkappa b. \qquad (2)$$

The correctness of the foregoing premise is confirmed by the course of the dependence of U_d(p). According to [9], U_d increases sharply when the interelectrode distance is being

Table 4

Gas	He	Ne	Ne	A
Cathode	Cylindrical	Flat	Cylindrical	Cylindrical
Current density on cathode, ma/cm^2	4	1.4	4	4
d, mm	0.77	0.99	0.96	~1.1

filled by the cathode parts. It is seen from Fig. 7 that the sharp increase in U_d begins at a pressure close to p_{max}.

Using the data given in [9], we can estimate the length of the cathode region d in different gases at $p = p_{max}$, as is done in Table 4. The relationships for argon in the region of low pressures were not determined, and consequently the value of p_{max} was not established accurately.

As can be seen from Table 4, the length of the cathode region d is practically the same for all three gases. It should be noted that, as in [9], d is the distance from the cathode to the boundary of the negative glow, where the concentration of the electrons is maximum. The total length of the cathode region exceeds this value. Recognizing that in our case $b = 10$ mm and $\varkappa = 0.48$, we obtain $\eta \simeq 0.2$.

As the height of the waveguide decreases, the optimum filling with the cathode region should be observed at a lower extent of it, i.e., at greater pressures. Such a character of the dependence of p_{max} on the loss curves and of the phase shift on the height of the waveguide is axially observed (see Tables 1 — 3). However, whereas the variation of p_{max} in the case of losses is, in accordance with (1), approximately inversely proportional to the height of the waveguide, p_{max} on the phase-shift curve varies to a considerably lesser degree.

An explanation of this difference necessitates more detailed information on the structure of the cathode region than is available at the present time. The effect described can be used in practice to reduce the attenuation introduced by a gas-discharge phase shifter, or the phase shift produced by a gas-discharge attenuator.

Apparently, to obtain maximum attenuations it is necessary to combine optimal filling of the waveguide by the cathode region with the condition

$$\nu_{col} = \nu^0_{col} p = \omega. \tag{3}$$

Substituting (1) in (2) we obtain

$$\eta = \frac{B}{p\varkappa b}. \tag{4}$$

Since $\omega = \pi c/\alpha a$, we have from (3) and (4)

$$\eta = \frac{\alpha B \nu^0_{col} a}{\pi c \varkappa b},$$

where a — with the waveguide; $\alpha = \lambda/\lambda_{cr}$; λ — wavelength; λ_{cr} — critical wavelength; c — velocity of light; ν^0_{col} — collision frequency at p = 1 mm Hg.

CONCLUSION

We have established that the phase shift and the attenuation, produced by the cathode parts of the glow discharge in a rectangular waveguide disclose a sharp maximum as functions of the pressure, in the pressure region where $\nu_{col} < \omega$. The presence of this maximum can be explained by the contraction of the cathode region towards the cathode, as a result of which the efficiency of interaction between the microwave oscillations and the plasma is reduced at pressures greater than p_{max}. This explanation is confirmed by the course of the dependence of the voltage of the discharge on the pressure and the character of the displacement of the attenuation maximum as the distance from the cathode to the anode is changed (long wall of the waveguide).

The maximum on the phase-shift curve shifts with changing height of the waveguide to a lesser extent than the maximum of the attenuation. This circumstance can be used to reduce the attenuation introduced by a gas-discharge phase shifter, or to reduce the phase shift produced by a gas-discharge attenuator.

Neon can be used to fill a phase shifter. The optimum values of θ and L obtained with it are close to those given in reference [2] where krypton was used ($0 = 180°$, $L \leq 0.3$ db). The presence of the minimum observed in neon on the $F(p)$ curve can be used in some cases to increase the working pressure. Helium should be used in gas-discharge attenuators of the absorbing type.

In conclusion I am grateful to E.M. Grekova who carried out the technological part of this investigation, and to V.E. Golant for a discussion of the results.

REFERENCES

1. R.M. Geiger, P.E. Dorney. IRE Convention Record, 1956, p. 5, 193.
2. D.H. Pringle, J. Scient. Instrum., 1955, 32, 4, 125.
3. E. Bradly, D. Pringle. J. Brit. Instn. Radio Engrs., 1955, 15, 1, 11.
4. D. Pringle, E. Bradly. J. Electronics, 1956, 1, 4, 389.
5. L. Goldstein, et al. Electr. Commun., 1952, 29, 3, 243.
6. F. Arams, Electronics, 1954, 27, 11, 168.
7. V.E. Golant, A.P. Zhilinskiy. ZhTF, Journal of Technical Physics, 1960, 30, 1, 15.
8. A.M. Starik. Radiotekhnika i elektronika, 1960, 5, 7, 1124.
9. A. von Engel and M. Steenbeck. Electrical Gas Discharges, Their Physics and Their Technology (in German), Edwards Brothers, Ann Arbor, Michigan, 1944.
10. B.J. Udelson, et al. J. Appl. Phys., 1957, 28, 6, 717.

Received by editors 28 November 1960

HOLLOW ELECTRON BEAM OF CONICAL SHAPE IN A "CONICAL" MAGNETIC FIELD

T. F. Loginova

Calculations and experimental data are given for a "conical" magnetic field, i.e., a field whose force lines are directed along the generatrices of coaxial cones with common vertex. The feasibility of focusing over a long stretch of a conical hollow electron beam with the aid of a conical magnetic field is experimentally demonstrated.

Many domestic and foreign papers have been published on the focusing of extensive electron beams of the cylindrical type. In some cases the need arises for electron beams of more complicated form. Papers devoted to the creation of such beams are much less frequently encountered. Thus, for example, we know of only one article [1] in which the question of shaping a conical electron beam with the aid of a magnetic field is discussed. It is shown theoretically in this paper that with the aid of a magnetic field of special form it is possible to compensate for the repulsion action of the space charge in a solid conical electron beam with small convergence angle, and to obtain a beam of the Brillouin type. The article contains no experimental information.

The present paper is devoted to the question of the focusing of a hollow electron beam, converging or diverging along a cone with definite angle of inclination to the axis (Fig. 1). Such an electron beam can be maintained over a great length, for example, with the aid of an axially symmetrical magnetic field, the force lines of which coincide over the entire length with the trajectories of the electrons. When the magnetic field is sufficiently large, the electrons deflected for some reason or another from its straight-line paths will twist around the force lines of the field in helical paths with small radii. When the magnetic field is increased without limit, the electrons will tend to move rigorously along the force lines, forming a so-called magnetically bounded beam [2] of conical shape.

The problem consequently consists above all in producing a magnetic field of the necessary configuration.

As is well known [3], the radial component R(r, z) and the longitudinal component Z(r, z) of the intensity vector in an axially symmetrical magnetic field can be expressed in terms of the intensity of the magnetic field on the axis, H(z), and its derivatives in the following fashion:

$$\begin{gathered} R\,(r,\,z) = -\frac{r}{2}H'(z) + \frac{r^3}{16}H'''(z) - \frac{r^5}{384}H^{(\mathrm{V})}(z) + \ldots \\ \ldots - \frac{(-1)^n}{n!\,(n+1)!}H^{(2n+1)}\,(z)\left(\frac{r}{2}\right)^{2n+1} + \ldots, \\ Z\,(r,\,z) = H\,(z) - \frac{r^2}{4}H''\,(z) + \\ + \frac{r^4}{64}H^{(\mathrm{IV})}(z) - \ldots + \frac{(-1)^n}{(n!)^2}H^{(2n)}\,(z)\left(\frac{r}{2}\right)^{2n} + \ldots . \end{gathered} \tag{1}$$

In the case of a slowly varying field or for small radii r, we can confine ourselves to the first terms of the series; then

$$\begin{aligned} R\,(r,\,z) &\simeq -\frac{r}{2}H'\,(z), \\ Z\,(r,\,z) &\simeq H\,(z). \end{aligned} \tag{2}$$

The inclination of the intensity vector at the point (r, z) to the axis is determined by the relation

$$\operatorname{tg}\alpha = \frac{R(r, z)}{Z(r, z)} = -\frac{r}{2}\frac{H'(z)}{H(z)}. \tag{3}$$

The condition under which the force lines of the magnetic field coincides at each point with the trajectory of the electron will be that the tangents to these lines coincide at each point, i.e.,

$$-\frac{r}{2}\frac{H'(z)}{H(z)} = \frac{dr}{dz}, \tag{4}$$

where $r(z)$ is the equation of the electron trajectory.

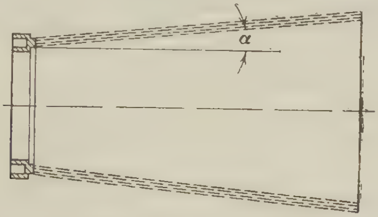


Fig. 1. Tubular electric beam of conical form

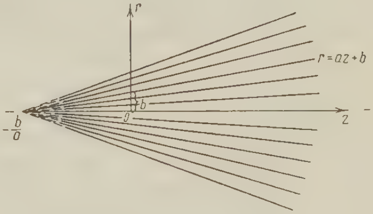


Fig. 2. Force lines of conical magnetic field

For a tubular thin-wall conical beam we have

$$r(z) = az + b,$$

where

$$a = \operatorname{tg}\alpha;\ b = r(z = 0).$$

In this case we obtain a very simple differential equation for $H(z)$:

$$\frac{dH(z)}{H(z)} = -\frac{2a\,dz}{az + b}, \tag{5}$$

Hence

$$H(z) = \frac{c}{\left(z + \frac{b}{a}\right)^2}, \tag{6}$$

where c is an integration constant, which depends on the absolute value of the specified magnetic field.

It is seen from (6) that in a magnetic field the force lines of which are directed along the generatrices of a cone defined by the equation $r = az + b$, the extent of the field along the axis should vary hyperbolically $1/(z + b/a)^2$.

In such a conical field, all the force lines coincide with the generatrices of coaxial round cones with a common vertex at the point $z = -b/a$. The generatrices of the cone are determined by the equation $r = k(az + b)$, where k is a parameter.

The picture of the conical field is shown in Fig. 2. Physically it can be realized only on a limited section along the z axis, since $H(z) \to \infty$ when $z \to -b/a$.

It must be noted that the solution (6) is not approximate but exact for the most general case. Indeed, upon substitution of (6) in expression (1), which contains no assumptions concerning the smallness of the radius or the slow variation of $H(z)$, we obtain for $r = az + b$, i.e., on the generatrices of the cone $r = az + b$,

$$R(r,z)=\frac{ac}{\left(z+\frac{b}{a}\right)^2}\sum_{n=0}^{\infty}\frac{(-1)^n(2n+1)!}{(n!)^2}\left(\frac{a}{2}\right)^{2n},$$
$$Z(r,z)=\frac{c}{\left(z+\frac{b}{a}\right)^2}\sum_{n=0}^{\infty}\frac{(-1)^n(2n+1)!}{(n!)^2}\left(\frac{a}{2}\right)^{2n}, \tag{7}$$

This means that $R(r, z)/Z(r, z) = a$, the vector $H(r, z)$ is directed along the generatrix at any point on the cone, and the limitations imposed previously to simplify the solutions are not necessary.

To check the possibility of focusing a conical electron beam we have chosen in our experiments the following parameters: $\alpha = 1°12'$, $\tan\alpha = a = 0.021$; $b = 0.9$ cm; $b/a = 43$.

The H-curve calculated for this case by the formula $H(z)/H_0 = 1850/(z+43)^2$ (we use the term H curve for the dependence of the intensity of the magnetic field at the axis of the focusing system on the distance along the axis) is shown in Fig. 3 by the solid line 2.

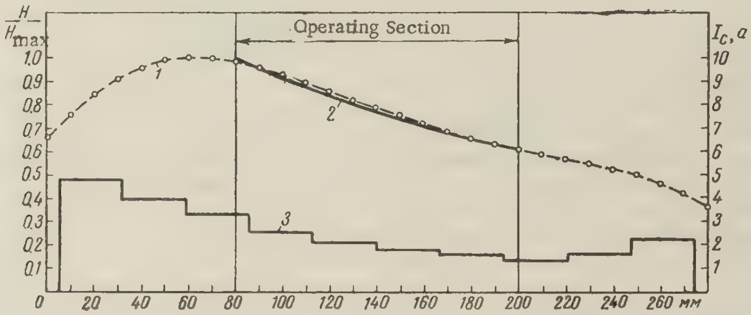


Fig. 3. Distribution of field intensity along the axis of a conical magnetic field:
1 — Experimental H-curve; 2 — calculated H-curve; 3 — distribution of current in sections of the solenoid.

The form of the magnetic field was chosen experimentally with the aid of a sectionalized solenoid, in which each of ten sections was fed separately. The current in each section could be regulated independently.

At a certain chosen distribution of the current in the sections, the H-curve of the solenoid was plotted, and after comparing the plotted curve with the theoretical one the distribution of the currents in the sections was corrected. The result was a chosen solenoid H-curve, differing from the theoretical one by more than 2% over a length of 140 mm (curve 1 on Fig. 3).

The distribution of the currents in the sections is shown for this case on curve 3 (Fig. 3).

The focusing of the conical electron beam in a conical magnetic field was checked experimentally with a model shown schematically in Fig. 4. The electron source was a Pierce-type gun with an annular cathode 0.6 mm wide. The end of the cathode and the focusing electrodes were tilted to the system axis at an angle of 1°12'.

The conical transit channel had a width $\Delta r = 2$ mm and a length $l = 100$ mm. The outer shell was solid along the entire length of the channel, and the internal rod was subdivided into two insulated parts with individual leads.

The tests have shown that when all sections of the solenoid are fed the same current (uniforn magnetic field) the electron beam did not reach the collector if the symmetry axis of the model coincided with the symmetry axis of the magnetic field. When the model was tilted relative to the solenoid axis by approximately 1°12', i.e., and one of the generatrices of the conical transit channel was aligned with the direction of the magnetic field, the maximum passage of the electron beam amounted to approximately 15%.

In a conical magnetic field having the same average value as a uniform field with the model placed on the axis of a solenoid, about 70 — 80% of the total current reached the collector. The comparative data for the operating conditions $U_a = 80$ v, $U_{ext} = U_{B_1} = U_{B_2} = U_{col} = 500$ v, are listed in the table (see following page).

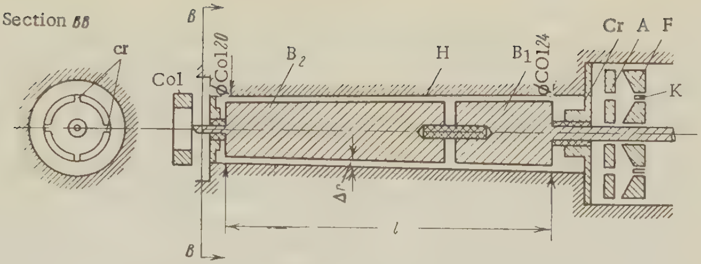


Fig. 4. Diagram of experimental model for verification of the focusing of a conical electron beam:

K — cathode; F — focusing electrode; A — anode; Cr — cross piece of centering rings; B_1, B_2 — sections of internal rod; H — outer shell; Col — collector; Δr — width of transit channel; l — length of transit channel.

The current settled essentially on the cross pieces of the centering rings (marked Cr) in Fig. 4), which were in contact with the outer shell of the transit channel H and which occupied approximately 12% of the cross sectional area of the channel.

Experience has shown that the currents settling on parts B_1 and B_2 of the internal rod are proportional to their lengths, this being evidence of the uniform distribution of the settling current over the length of the transit channel.

As the collector current is increased to 100 ma, the percentage current passing through the collector remains practically unchanged at a fixed magnetic field (Fig. 5).

We checked the dependence of the focusing on the absolute value of the magnetic field: when the maximum field intensity is reduced from 1200 to 780 gauss, i.e., by 1.5 times, the percentage of the current transmitted changes from 80 to 75%, i.e., merely 5%.

Magnetic field	Position of model in the solenoid	I_{ext}, ma	I_{B1}, ma	I_{B2}, ma	I_{col}, ma	ΣI, ma	ϰ, %
Uniform	along the axis	6.2	0	0	0	6.2	0
	at an angle to the axis	5.2	0.05	0.2	1.0	6.45	15.5
Conical	at an angle to the axis	3.8	0.6	0.9	1.3	6.6	20
	along the axis	1.1	0.15	0.25	5.25	6.75	78

Note. I_{ext} — current settling on the external shell; I_{B_1} and I_{B_2} — currents settling on parts B_1 and B_2 of the internal rod, respectively; I_{col} — collector current; ΣI — summary current in the electron beam; ϰ — percentage of current reaching the collector.

The tests have also shown that a conical beam is not sensitive to a slight change in the shape of the magnetic field. Thus, a 10% change in the current of one of ten sections, relative to the theoretical value, caused no noticeable change in the beam passage.

The second experiment on focusing a conical electron beam was carried out with a model analogous in construction to that shown in Fig. 4, but with a transit channel of width Δr = 1 mm and length l = 120 mm.

In this model, placed in a conical magnetic field having intensity drop from 1100 to 1900 gauss over the working portion, the electron current reaching the collector was up to 200 ma, with a passage percentage on the order of 65-70.

Traces of heating on the electrodes demonstrated that essentially the settling of the electron current was not over the length of the transit channel, but on the diaphragm of the second anode, which was in contact with the outer

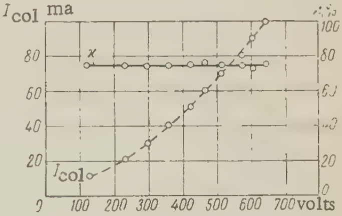


Fig. 5. Dependence of passage of collector current on the anode voltage in a conical beam; I_{col} — collector current; ϰ — percentage of current passing to collector

electrode of the transit channel. This means that the settling was due not to defects in magnetic focusing, but to poor quality of the electron gun, which produced too wide a beam at the output.

Measurements have shown that, other conditions being fixed, the focusing is practically independent of the potential of the transit channel within a wide range of variations (from 500 to 4,000 v).

It was also found that shifting the model along the solenoid axis by 30 — 40 mm relative to the calculated position changes the transmission by not more than 3 — 4%.

Thus, the experiments carried out have shown that with the aid of a conical magnetic field it is possible to focus a conical hollow electron beam of small thickness over a length more than 100 mm. In our investigations we used a beam with small convergence angle and with relative low current density, on the order of 0.1 — 0.4 amp/cm^2. However, with the aid of a conical magnetic field it is possible to effect focusing in solid convergent beams with different convergence angles. In the case of large convergence angles, a rapid variation of field intensity along the axis is necessary in order to produce the conical magnetic field. This variation can be produced only with the aid of magnetic shields and end pieces.

To focus thick conical beams with large current density, the magnetic fields must have greater absolute values.

In conclusion, I consider it my pleasant duty to thank L. P. Lisovskiy for help in the work and for advice.

REFERENCES

1. M.E. Hines, Proc. IRE, 1952, 40, 1, 61.
2. J.R. Pierce. Theory and Design of Electron Beams, Van Nostrand, New York, 1949. (Russian translation, p. 162).
3. A. Rusterholz. Elektronenoptik, Birkhauser, Basel, 1950 (Russian translation, p. 67).

Received by editor 3 December 1960

CONTRIBUTION TO THE THEORY OF A TRAVELING WAVE QUANTUM PARAMAGNETIC AMPLIFIER

G.S. Misezhnikov, V.B. Shteynshleyger

The dispersion characteristic of a slow-wave system of the pin-comb type, used in a quantum paramagnetic amplifier, is calculated; an approximate expression is given for the field in such a system; the coefficient of utilization of the magnetic field, which is contained in the expression for the gain of the quantum paramagnetic traveling wave amplifier, is calculated for the forward and backward waves.

INTRODUCTION

The gain G (in decibels) of a quantum paramagnetic traveling-wave amplifier is given by the expression [1]

$$G = 4.35\,|\chi''_{\max}|\,F\frac{\omega l}{v_g},$$

where $\chi''_{\max}$ is the maximum value of the magnetic susceptibility of the paramagnetic substance (in amplification $\chi''_{\max} < 0$); ω — circular frequency; l — active length of slow-wave system; v_g — group velocity of wave in the slow-wave system; F — "filling coefficient," which determines the filling of the cross section of the system with the paramagnetic substance and the efficiency of interaction of the latter with the magnetic field in the system (the quantity F should have been called the "coefficient of utilization of the magnetic field" which is closer to the physics of the phenomenon).

For a theoretical determination of G it is necessary thus to know not only the material parameter $\chi''_{\max}$, but also the quantity v_g and F.

The value of v_g can be determined from the dispersion characteristic of the slow-wave system* (being the reciprocal of the derivative of the phase constant with respect to the frequency); it is, therefore, necessary first to calculate the dispersion characteristic of the system.

To calculate F, it is necessary to determine the magnetic field in the slow-wave system.

In the present paper we give the corresponding solutions for the case of a slow-wave system of the pin-comb type in a waveguide, as used in [1]. Since it is desirable to make the comb as thin as possible, so as to increase the coefficient F, we shall consider an idealized slow-wave system in the form of an infinitesimally thin pin-comb in a rectangular waveguide (see Fig. 1). For such a slow-wave system model, the known methods of calculation with the aid of a modified theory of long lines [1] leads to large errors, since the field at the pin is no longer screened from the field in the remaining part of the system by the neighboring pins, nor are there any accurate methods for taking into account the stray field at the ends of the pins.

We develop below an electrodynamic method for the calculation of a pin-type slow-wave system for a traveling-wave quantum paramagnetic amplifier, a method free of these shortcomings and yielding a solution with any degree of accuracy. We assume here that the space above and below the comb is completely filled with a homogeneous dielectric. This enables

---

*We speak here of a theoretical (computational) determination of v_g. For an experimental determination of v_g there is no need for finding the dispersion characteristic (see, for example [2]).

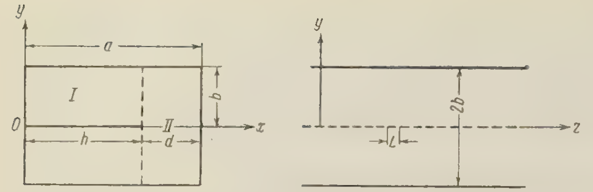


Fig. 1. Slow-wave system of the pin-comb type in a waveguide. Transverse and longitudinal cross sections.

us to regard the system as empty (inasmuch as a full homogeneous filling yields simply a similarity change in dimensions) and avoid computational complications that arise in the case of partial filling with the dielectric. The main features of the apparatus are not lost sight of in this case. The determination of the dispersion equation will be based on the method of joining the fields at the boundaries of the elementary regions, into which the section of the waveguide can be subdivided.

1. DERIVATION OF DISPERSION EQUATION

The slow-wave system under consideration is shown in Fig. 1. The wave propagates along the z axis. The pins, infinitesimally thin in the y direction, are located in the plane $y = 0$. It is assumed that the period of the system L is sufficiently small compared with the wavelength in the system Λ, so that we need retain in the expansion of the field in the spatial harmonics only the fundamental spatial harmonic, and consider the pin comb as an anisotropically conducting plane, having ideal conductivity in the x direction (along the pins) and non-conducting in the perpendicular direction.

The boundary conditions in the plane of the comb, corresponding to the assumption of anisotropic conductivity, have the form

$$E_x|_{y=0} = 0\,, \tag{1}$$

$$H_x|_{y=+0} = H_x|_{y=-0}\,. \tag{2}$$

The last equation denotes the absence of surface currents in the z direction.

The boundary conditions (1) and (2) jointly with the boundary conditions on the walls of the waveguide enable us to determine the field of the fundamental spatial harmonic and find the dispersion characteristic of the system.

The components of the electric and magnetic fields can be obtained from expressions for the electric and magnetic Hertz functions, U and V respectively, with the aid of the following equation [3]:

$$\begin{aligned} E_x &= -j\left(\beta\frac{\partial U}{\partial x} + k\frac{\partial V}{\partial y}\right), \\ E_y &= -j\left(\beta\frac{\partial U}{\partial y} - k\frac{\partial V}{\partial x}\right), \\ E_z &= (k^2-\beta^2)\,U, \end{aligned} \tag{3}$$

$$\begin{aligned} H_x &= -j\sqrt{\frac{\varepsilon}{\mu}}\left(\beta\frac{\partial V}{\partial x} - k\frac{\partial U}{\partial y}\right), \\ H_y &= -j\sqrt{\frac{\varepsilon}{\mu}}\left(\beta\frac{\partial V}{\partial y} + k\frac{\partial U}{\partial x}\right), \\ H_z &= \sqrt{\frac{\varepsilon}{\mu}}\,(k^2-\beta^2)\,V, \end{aligned} \tag{4}$$

where ϵ and μ are the dielectric constant and the permeability of the medium; k is the phase constant of the wave in free space; β is the phase constant of the wave in the system.

The factor $\exp j(\omega t - \beta z)$, which is common to all vectors, will be omitted henceforth. The functions U and V satisfy the wave equation

$$\begin{aligned} \Delta U + k^2 U &= 0, \\ V + k^2 V &= 0. \end{aligned} \tag{5}$$

Let us separate in the cross section of the slow-wave system two rectangular regions (see Fig. 1), in each of which the boundary conditions are separable in the variables x and y:

region I — between the comb and the wide wall of the waveguide, region II — between the comb and the narrow wall of the waveguide.

Let us set up a solution for U and V in the form of functions (partial waves) that satisfy the wave equation and the boundary condition $E_{tan} = 0$ on the waveguide walls. The field in region I is determined essentially by the TEM wave propagating along the x axis (in this wave $E_x = H_x = 0$), and higher modes exist only at the ends of the pins. In order for this wave to satisfy the boundary conditions on the walls of the waveguide ($E_y = 0$ for $x = 0$; $E_z = 0$ for $x = 0$ and $y = b$), the Hertz function of this wave must be written in the following form:

$$\begin{aligned} U_{TEM} &= U_{01} \operatorname{sh} \beta (b - y) \sin kx, \\ V_{TEM} &= V_{01} \operatorname{ch} \beta (b - y) \cos kx, \end{aligned}$$

and since the E_x component vanishes everywhere, we have

$$U_{01} = V_{01}.$$

The remaining partial waves are represented in such a way as to satisfy, in addition to the aforementioned boundary conditions, also the boundary condition $E_x = 0$ for $y = 0$ (see condition (1)) and $y = b$.

Thus, (when $y > 0$)

$$\begin{aligned} U_I &= U_{01} \operatorname{sh} \beta (b - y) \sin kx + \sum_{m=1}^{\infty} U_{m1} \operatorname{sh} \gamma_{m1} x \sin \alpha_{m1} y, \\ V_I &= U_{01} \operatorname{ch} \beta (b - y) \cos kx + \sum_{m=1}^{\infty} V_{m1} \operatorname{ch} \gamma_{m1} x \cos \alpha_{m1} y, \end{aligned} \tag{6}$$

where

$$\alpha_{m1} = \frac{m\pi}{b}\,; \gamma_{m1} = \sqrt{\beta^2 - k^2 + \alpha_{m1}^2}.$$

The solution contains an unknown phase constant β and unknown amplitudes U_{m1} and V_{m1} of the partial waves. As a result of the symmetry of the field relative to the plane of the comb, E_z is an even function, and H_z is an odd function of the coordinate y. In this connection (see (4)), H_x is an odd function of y. Therefore the boundary condition (2) is satisfied if

$$H_x |_{y=0} = 0; \tag{2a}$$

Substituting the expressions U_I and V_I from (6) into condition (2a) we obtain

$$\sum_{m=1}^{\infty} (\beta\gamma_{m1} V_{m1} - k\alpha_{m1} U_{m1}) \operatorname{sh} \gamma_{m1} x = 0 \tag{7}$$

or, integrating (7) with respect to x from 0 to h,

$$\sum_{m=1}^{\infty} (\beta\gamma_{m1} V_{m1} - k\alpha_{m1} U_{m1}) (\operatorname{ch} \gamma_{m1} h - 1) = 0. \tag{8}$$

In region II, the electric and magnetic fields can be represented in the form of superpositions of TE and TM modes (satisfying the following boundary conditions on the waveguide wall: $E_x = 0$ for $y = \pm b$; $E_y = 0$ for $x = a$; $E_z = 0$ for $y = \pm b$ and $x = a$)

$$U_{II} = \sum_{m=1}^{\infty} U_{m2} \operatorname{sh} \gamma_{m2} (a - x) \cos \alpha_{m2} y,$$

$$V_{II}=\sum_{m-1}^{\infty} V_{m2}\,\text{ch}\,\gamma_{m2}(a-x)\sin\alpha_{m2}y, \tag{9}$$

where $\alpha_{m2}=(2m-1)\pi/2b;\ \gamma_{m2}=\sqrt{\beta^2-k^2+\alpha_{m2}^2};$ U_{m2} and V_{m2} are the unknown amplitudes of the partial waves in region II.

The equality of the tangential components of the electric and magnetic fields on the boundary of regions I and II when $x = 0$ leads to the following system of equations

$$U_{01}\,\text{sh}\,\beta(b-y)\sin kh+\sum_{r=1}^{\infty}U_{r1}\,\text{sh}\,\gamma_{r1}h\sin\alpha_{r1}y=$$

$$=\sum_{m=1}^{\infty}U_{m2}\,\text{sh}\,\gamma_{m2}d\cos\alpha_{m2}y, \tag{10a}$$

$$U_{01}\,\text{ch}\,\beta(b-y)\cos kh+\sum_{r=1}^{\infty}V_{r1}\,\text{ch}\,\gamma_{r1}h\cos\alpha_{r1}y=$$

$$=\sum_{m=1}^{\infty}V_{m2}\,\text{ch}\,\gamma_{m2}d\sin\alpha_{m2}y, \tag{10b}$$

$$kU_{01}\,\text{sh}\beta(b-y)\cos kh+\sum_{r=1}^{\infty}U_{r1}\gamma_{r1}\,\text{ch}\,\gamma_{r1}h\sin\alpha_{r1}y=$$

$$=-\sum_{m=1}^{\infty}U_{m2}\gamma_{m2}\,\text{ch}\,\gamma_{m2}d\cos\alpha_{m2}y, \tag{10c}$$

$$kU_{01}\,\text{ch}\,\beta(b-y)\sin kh-\sum_{r=1}^{\infty}V_{r1}\gamma_{r1}\,\text{sh}\,\gamma_{r1}h\cos\alpha_{r1}y=$$

$$=\sum_{m=1}^{\infty}V_{m2}\gamma_{m2}\,\text{sh}\,\gamma_{m2}d\sin\alpha_{m2}y. \tag{10d}$$

Multiplying (10a) and (10d) by $\cos\alpha_{m2}y$ and $\sin\alpha_{m2}y$ respectively, and integrating with respect to y from 0 to b, we obtain

$$U_{m2}=\frac{2}{b}\times$$

$$\times\frac{U_{01}\sin kh\int\limits_0^b\text{sh}\,\beta(b-y)\cos\alpha_{m2}y\,dy+\sum\limits_{r=1}^{\infty}U_{r1}\,\text{sh}\,\gamma_{r1}h\int\limits_0^b\sin\alpha_{r1}y\cos\alpha_{m2}y\,dy}{\text{sh}\,\gamma_{m2}d}, \tag{11a}$$

$$V_{m2}=\frac{2}{b}\times$$

$$\times\frac{kU_{01}\sin kh\int\limits_0^b\text{ch}\,\beta(b-y)\sin\alpha_{m2}y\,dy-\sum\limits_{r=1}^{\infty}V_{r1}\gamma_{r1}\,\text{sh}\,\gamma_{r1}h\int\limits_0^b\cos\alpha_{r1}y\sin\alpha_{m2}dy}{\gamma_{m2}\,\text{sh}\,\gamma_{m2}d}. \tag{11b}$$

Multiplying (10b) and (10c) respectively by $\cos\alpha_{n1}y$ and $\sin\alpha_{n1}y$, integrating with respect to y from 0 to b, and integrating U_{m2} and V_{m2} from (11a) and (11b) into the resultant equations, we obtain

$$U_{01}\left(\cos kh\int\limits_0^b\text{ch}\,\beta(b-y)\cos\alpha_{n1}y\,dy-A_nk\sin kh\right)+$$

$$+\sum_{r=1}^{\infty}V_{r1}\left(B_{nr}\gamma_{r1}\,\text{sh}\,\gamma_{r1}h+\delta_{nr}\frac{b}{2}\,\text{ch}\,\gamma_{r1}h\right)=0,$$

$$U_{01}\left(k\cos kh\int_0^b \operatorname{sh}\beta(b-y)\sin\alpha_{n1}y\,dy + C_n\sin kh\right) + \\ + \sum_{r=1}^{\infty} U_{r1}\left(D_{nr}\operatorname{sh}\gamma_{r1}h + \delta_{nr}\frac{b}{2}\gamma_{r1}\operatorname{ch}\gamma_{r1}h\right) = 0, \tag{12}$$

where

$$A_n = \frac{2}{b}\sum_{m=1}^{\infty}\frac{\int_0^b \operatorname{ch}\beta(b-y)\sin\alpha_{m2}y\,dy\int_0^b \sin\alpha_{m2}y\cos\alpha_{n1}y\,dy}{\gamma_{m2}\operatorname{th}\gamma_{m2}d}; \tag{13}$$

$$B_{nr} = \frac{2}{b}\sum_{m=1}^{\infty}\frac{\int_0^b \cos\alpha_{r1}y\sin\alpha_{m2}y\,dy\int_0^b \sin\alpha_{m2}y\cos\alpha_{n1}y\,dy}{\gamma_{m2}\operatorname{th}\gamma_{m2}d}; \tag{14}$$

$$C_n = \frac{2}{b}\sum_{m=1}^{\infty}\frac{\int_0^b \operatorname{sh}\beta(b-y)\cos\alpha_{m2}y\,dy\int_0^b \cos\alpha_{m2}y\sin\alpha_{n1}y\,dy}{\dfrac{\operatorname{th}\gamma_{m2}d}{\gamma_{m2}}}; \tag{15}$$

$$D_{nr} = \frac{2}{b}\sum_{m=1}^{\infty}\frac{\int_0^b \sin\alpha_{r1}y\cos\alpha_{m2}y\,dy\int_0^b \cos\alpha_{m2}y\sin\alpha_{n1}y\,dy}{\dfrac{\operatorname{th}\gamma_{m2}d}{\gamma_{m2}}}; \tag{16}$$

$$\delta_{nr} = \begin{cases} 1 \text{ when } n = r, \\ 0 \text{ when } n \neq r. \end{cases}$$

Forming in accordance with (12) a system of equations, each of which corresponds to a certain value of n (n = 1, 2, . . .) and contains an infinite number of terms (r = 1, 2, . . .), and adding (8) to them, we obtain an infinite system of linear homogeneous equations, which has nontrivial solutions if its determinant vanishes, i.e.,

$$\begin{vmatrix} a_1 & b_{11}\ldots b_{1r} & 0\ldots 0 \\ \cdots & \cdots & \cdots \\ a_n & b_{n1}\ldots b_{nr} & 0\ldots 0 \\ c_1 & 0\ldots 0 & d_{11}\ldots d_{1r} \\ \cdots & \cdots & \cdots \\ c_n & 0\ldots 0 & d_{n1}\ldots d_{nr} \\ 0 & f_1\ldots f_r & g_1\ldots g_r \end{vmatrix} = 0, \tag{17}$$

where

$$a_n = \cos kh\int_0^b \operatorname{ch}\beta(b-y)\cos\alpha_{n1}y\,dy - A_n k\sin kh;$$

$$b_{nr} = B_{nr}\gamma_{r1}\operatorname{sh}\gamma_{r1}h + \delta_{nr}\frac{b}{2}\operatorname{ch}\gamma_{r1}h;$$

$$c_n = k\cos kh\int_0^b \operatorname{sh}\beta(b-y)\sin\alpha_{n1}y\,dy + C_n\sin kh;$$

$$d_{nr} = D_{nr}\operatorname{sh}\gamma_{r1}h + \delta_{nr}\frac{b}{2}\gamma_{r1}\operatorname{ch}\gamma_{r1}h;$$

$$f_r = \beta\gamma_{r1}(\operatorname{ch}\gamma_{r1}h - 1); \quad g_r = k\alpha_{r1}(\operatorname{ch}\gamma_{r1}h - 1).$$

2. RESULTS OF CALCULATION OF THE DISPERSION CHARACTERISTIC

The dispersion equation (17) derived in Section 1 enables us to find the dependence of β on k for given geometric dimensions of the system, with any desired degree of accuracy. For this purpose we approximate the infinite determinant with determinants containing 3 x 3, 5 x 5, etc. terms, depending on the required computational accuracy.

We obtained a zero-approximation solution of clearest form, amenable to physical interpretation, by confining ourselves to the fundamental TEM wave in the description of the field in region I.

For this purpose we put $n = 0$ in the system (12), thereby reducing it to the single equation

$$U_{01}\left(\cos kh \int_0^b \operatorname{ch} \beta (b - y)\, dy - A_0 k \sin kh\right) = 0 \tag{18}$$

(This condition corresponds to joining the fundamental field components E_y and H_z on the boundary of the regions, where $x = h$, neglecting higher types of modes in region I).

From (18), using the expression (13) for A_0, we obtain

$$\frac{1}{k \operatorname{tg} kh} = \frac{2}{b} \frac{\beta}{\operatorname{th} \beta b} \sum_{m=1}^{\infty} \frac{1}{(\beta^2 + \alpha_{m2}^2)\, \gamma_{m2} \operatorname{th} \gamma_{m2} d}. \tag{19}$$

Corresponding to the boundaries of the passband of the slow-wave system are the values $\beta = 0$ and $\beta = \infty$. Thus, we obtain from (19)

$$\left.\frac{\nu}{4h}\right|_{\beta=\infty} = \nu_\infty' = 1, \tag{20}$$

$$\left.\frac{\nu}{4h}\right|_{\beta=0} \simeq \nu_0' = 1 + \frac{16\, b}{\pi^3 h} \sum_{m=1}^{\infty} \frac{1}{(2m-1)^3 \operatorname{th} \frac{(2m-1)\pi d}{2b}}. \tag{21}$$

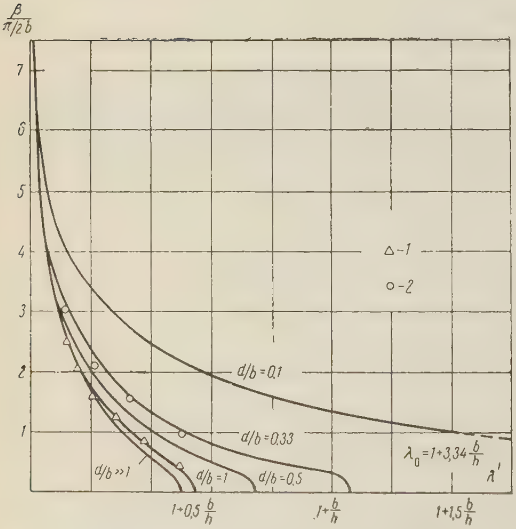


Fig. 2. Dispersion characteristic of slow-wave system (zeroth approximation for various d/b: 1 — experimental values obtained in [1] for d/b = 1; 2 — the same for d/b = 0.33.

The last equation is approximate and obtained under the approximation $kh \tan kh \simeq \simeq \frac{kh}{\frac{\pi}{2} - kh}$, which is valid when kh is close to $\pi/2$, i.e., when the slow-wave system has a small bandwidth (compared with ν) In particular, we have when $d/b \gg 1$

$$\nu_0' \simeq 1 + 0.5 \frac{b}{h}. \tag{22}$$

The dispersion characteristics of the slow-wave system, calculated by formula (19) for different values of d/b, are shown in Fig. 2. The same figure shows the experimental points obtained in [1] for two values of d/b (d/b = 1 and 0.33). It is easy to verify that experiment agrees well with the theory.

When the above-described slow-wave system is used in paramagnetic traveling-wave amplifiers, an increase of the slowing down in the group velocity $v_g = \frac{1}{\partial \beta / \partial \omega}$ is attained by narrowing down the pass band of the slow-wave system. For this purpose it is necessary to tend to increase the distance d between the pins and the narrow wall of the waveguide. As can be seen from Fig. 2, the value d/b = 1 is sufficient, and any further increase in d does not reduce the band. In this limiting case ($d/b \gg 1$) we have from (22)

$$\frac{\Delta \nu}{\nu} = 0.5 \frac{b}{h},$$

and the subsequent narrowing down of the band is reached by reducing the distance b from the comb to the broad wall of the waveguide.

Fig. 3 shows a plot characterizing the change in the coefficient of the slowing down of the group velocity within the pass-band of the slow-wave system, plotted by graphically differentiating the dispersion characteristic of Fig. 2 for the case $d/b \gg 1$.

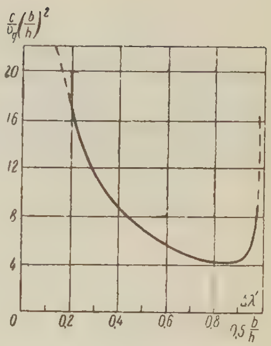


Fig. 3. Coefficient of group-velocity slowing down.

It is seen from Fig. 3 that within the limits of the pass-band (near the low-frequency edge of the band) we can choose a considerable portion with slowing-varying group velocity, whereas the phase velocity (see Fig. 2) changes by a factor of several times within the limits of the same portion.

3. COEFFICIENT OF UTILIZATION OF MAGNETIC FIELD

The coefficient of utilization F of the magnetic field, entering into the formula for the coefficient of amplification (see the introduction), can be expressed by the following formula, which follows from [4]:

$$F = \frac{\int_{\sigma_{\mathrm{cr}}} \left| \vec{H}^* \vec{S}_{pq} \right|^2 d\sigma}{\left(\vec{S}_{pq} \vec{S}^*_{pq} \right) \int_{\sigma} \left(\vec{H} \vec{H}^* \right) d\sigma}, \tag{23}$$

where (24)

$$\vec{S}_{pq} = \langle p | \hat{\vec{S}} | q \rangle$$

is (in the Dirac notation) the matrix element of the vector spin operator $\hat{\vec{S}}$ (with components $\hat{S}_x$, $\hat{S}_y$, $\hat{S}_z$) for the quantum transition, at the signal frequency, between energy levels p and q in the paramagnetic crystal.

The matrix elements of the operators $\hat{S}_x$, $\hat{S}_y$, and $\hat{S}_z$ for transitions in ruby crystal, used in a traveling-wave paramagnetic amplifier [1], are listed, for example, in the table of the review [5]; these data are based on the assumption that the z axis coincides with the trigonal (optical) axis of the crystal, and the constant magnetic field is in the plane xoz. In this case the matrix elements can be written in the form m_x, jm_y, m_z (the table lists the quantities $\alpha = 2m_x$, $\beta = 2m_y$, and $\gamma = 2m_z$), where m_x, m_y, and m_z are real numbers.

We shall approximate the variable high-frequency field $\vec{H}$ in the slow-wave system by a TEM wave described in terms of the first members of the expressions for U_I and V_I in formula (6).* If we write down the field components $\vec{H}$ from (6) and (4), we interchange the axes y and z, so as to retain the aforementioned conditions and form of writing down the matrix element; under such a change, the axes z' and x will lie in the plane of the transverse cross section of the waveguide, in which the constant magnetic field is also located (see [1]).

Taking this remark into account, we obtain from (6) and (4) the component along the pin

$$H_x = 0; \tag{25a}$$

The longitudinal component (along the waveguide axis)

$$H_{y'} = A \operatorname{ch} \beta (b - z'); \tag{25b}$$

the transverse component (perpendicular to the comb)

---

*We neglect the field perturbations due to the presence of paramagnetic ions.

$$H_{z'} = \mp jA \operatorname{sh} \beta (b - z'). \tag{25c}$$

Here A is a common factor, the primes denote the new axis, and the plus-and-minus signs correspond to waves propagating in opposite directions (along the new y' axis). A change in sign leads to different values of F for these waves.

Formulas (25) were written for $z' > 0$. To find the field for $z' < 0$, we can use the same formulas, bearing in mind here that $H_{y'}$, is an odd function and $H_{z'}$ is an even function of z'.

Using the notation for the matrix elements, we obtain from (23)

$$F = \frac{\int\limits_{\sigma_{\text{кр}}} \left| j m_y H^*_{y'} + m_z H^*_{z'} \right|^2 d\sigma}{(m_x^2 + m_y^2 + m_z^2) \int\limits_{\sigma} \left(\left| H_{y'} \right|^2 + \left| H_{z'} \right|^2 \right) d\sigma} \tag{26}$$

Substituting (25) in (26) and integrating with respect to z' from zero to b (we assume the filling complete along the x-axis, and neglect the portion $x > h$), we obtain

$$F_{\pm} = \frac{1}{2} \left(\frac{1}{2} \frac{m_y^2 + m_z^2}{m_x^2 + m_y^2 + m_z^2} - \frac{m_y^2 - m_z^2}{m_x^2 + m_y^2 + m_z^2} \frac{\beta b}{\operatorname{sh}^2 \beta b} \pm \frac{m_y m_z}{m_x^2 + m_y^2 + m_z^2} \operatorname{th} \beta b \right). \tag{27}$$

Here, as in (25), the two signs correspond to the propagation of the waves in opposite directions along the system. It is clear from (26) that to increase F (in the amplifying direction) it is desirable to use transitions such that $m_x^2 \ll m_y^2 + m_z^2$; for this case, neglecting the term m_x^2, we get

$$F_{\pm} = \frac{1}{4} \left(1 - \frac{1 - \eta^2}{1 + \eta^2} \frac{2\beta b}{\operatorname{sh} 2\beta b} \right) \pm \frac{\eta}{2(1 + \eta^2)} \operatorname{th} \beta b, \tag{27a}$$

where $\eta = m_y/m_z$ is the "coefficient of ellipticity of polarization of the quantum transition." The effect of directivity of the amplification is characterized by the ratio $R = F_+/F_-$, which we shall call the "coefficient of non-reciprocity of amplification".

Within the limits of the passband of the slow-wave system, the quantity βb varies from small values on the low-frequency edge of the band to a value $\beta b \gg 1$ on the high-frequency edge of the band. Accordingly, for a given value of η, we can determine from (27a) the variation of F within the band for both propagation directions.

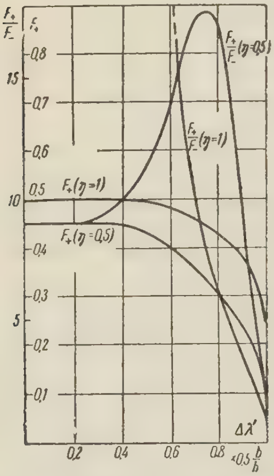


Fig. 4. Coefficient of field utilization and coefficient of non-reciprocity.

Fig. 4 shows the characteristics of F_+ in the amplification direction and of the coefficient $R = F_+/F_-$ within the pass band of the slow-wave system, for two values of the ellipticity coefficient η ($\eta = 1$ and $\eta = 0.5$).

It follows from Figs. 2, 3, and 4 that in spite of the appreciable change in β, and consequently also in the phase velocity within the portion of the band in which the group velocity changes little, the coefficient F changes insignificantly within this portion, i.e., it depends little on the phase velocity.

Consequently, the gain of an amplifier with the above-considered slow-wave system is also independent of the phase velocity, and is determined by the group-velocity slowing-down factor.

CONCLUSION

The relations obtained above enable us to calculate the main characteristics of a paramagnetic traveling-wave amplifier with a pin-type slow-wave system.

In some cases it may be necessary to take additional account of such factors as the incomplete filling of the volume with the dielectric, the finite thickness of the comb, and the presence of spatial field harmonics.

REFERENCES

1. R.W. DeGrasse, E.O. Schulz-DuBois, H.E.D. Scovil. Bell System Techn., J., 1959, 38, 305.
2. V.B. Shteinshleiger, G.S. Misezhnikov. Instruments and Experimental Techniques, 1959, 6, 133.
3. L.A. Vainshtein. Electromagnetic waves, Soviet Radio Press, 1957.
4. E.O. Schulz-DuBois, H.E.D. Scovil, R.W. DeGrasse. Bell System Techn. J., 1959, 38, 335.
5. J. Weber. Rev. Mod. Phys., 1959, 31, 3, 681.

Received by editors 22 February 1961

INVESTIGATION OF A TRANSISTORIZED LOW-FREQUENCY SELF-OSCILLATOR

A.S. Maydankovskiy

We consider a self-oscillator using an alloyed transistor. The calculations are based on characteristics deduced from the electronics of the transistor. In one case expressions are derived for the frequency of the self-oscillations and for the optimal load. The effect of the base resistance is investigated. The results of the experiment confirm the deduction of the theory.

INTRODUCTION

An investigation of transistorized self-oscillators is of practical interest in view of the extensive use of transistors in modern radio engineering.

Several papers devoted to the nonlinear theory of semiconductor self-oscillators have already been published [1, 2, 3]. However, in all the papers the calculation is based on an approximation of the experimentally plotted transistor characteristics.

In the present paper we give an analysis of low-frequency self-oscillators using junction diode characteristics obtained from the electronics of the latter. The solution of the problem was obtained by the method of N.N. Bogolyubov [4], the second approximation being considered for the frequency.

We have investigated a self-oscillator with inductive feedback using an alloyed transistor with grounded based.

1. DERIVATION OF INITIAL EQUATIONS

The circuit diagram of the self-oscillator is shown in Fig. 1. For the sake of physical clarity, the capacitance of the collector junction C_c and the base resistance $r_b = r'_b + r''_b$ are shown in the external circuits [5, 6]; g_1 is the loss conductivity in the resonant circuit; the meaning of the remaining symbols used in the present paper is clear from Fig. 1.

The equation for the voltage across the circuit capacitor C_1 can be written in the form

$$L_1C_1\ddot{U} + L_1g_1\dot{U} + U - L_1\frac{di_c}{dt} + M\frac{di_e}{dt} + L_1\frac{di_{cc}}{dt} = 0. \tag{1}$$

The capacitive current of the collector junction i_{cc} is given by [7, 8, 9]

$$i_{cc} = C_{c0}\left(\frac{-V_{cc}}{E'_c}\right)^{-1/2}\dot{V}_{cc}. \tag{2}$$

Here C_{c0} is the capacitance of the collector junction in the absence of an alternating voltage on the collector; $E'_c = E_c + V_D$, where V_D is the diffusion potential of the junction.

We shall henceforth consider only frequency regions where the inequality $r''_b C_c \omega \ll 1$ is satisfied. With this, the voltage on the capacitor C_c and on the collector junction can be written in the form

$$V_{cc} = V_c \simeq U - E_c. \tag{3}$$

In the case of low frequencies, when transit phenomena in the base do not play any role, the dependence of the emitter and collector currents on the voltages at the transistor junctions can be obtained at large signal levels, with account taken of the base-width modulation effect, by solving the diffusion equation for the transistor [5, 6, 9, 10]. These have the following form

$$\begin{gathered} i_e = i_0\{(e^{aV_e} - 1) - \alpha_0(e^{aV_c} - 1)\}\left(1 - \frac{w_1}{w_0}\right), \\ i_c = i_0\{\alpha_0(e^{aV_e} - 1) - (e^{aV_c} - 1)\}\left(1 - \frac{w_1}{w_0}\right), \\ w_1 = -d_{c0}\sqrt{1 - \frac{U}{E'_c}}. \end{gathered} \tag{4}$$

Here i_0 is the saturation current of the collector; α_0 — the current gain of the grounded-base transistor; w_0 and d_{c0} — width of base and width of the collector junction in the absence of an ac voltage; w_1 — ac component of the base width, due to the modulation of the collector junction; $a = q/kT$; q — charge of electron; k — Boltzmann's constant; T — absolute temperature.

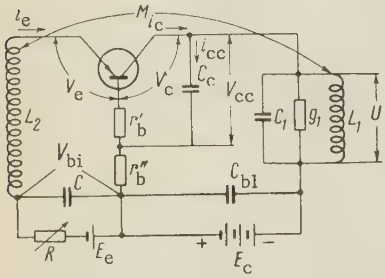


Fig. 1. Principal diagram of self-oscillator.

The emitter-junction voltage is related with the resonant-circuit voltage by the equation

$$V_e = k_1U - i_{cc}r''_b + V_{bi} - L_2(1 - K^2)\frac{di_e}{dt} - r_b(i_e - i_c), \tag{5}$$

where $K^2 = M^2/L_1L_2$; $k_1 = M/L_1$, and the bias voltage is determined from the relation

$$\dot{V}_{bi} + \frac{1}{\tau_{bi}}(V_{bi} - E_e) = -\frac{R}{\tau_{bi}}i_e;\ \tau_{bi} = RC. \tag{6}$$

For the sake of convenience, we divide the current i_{cc} into two components — the current i_{cc1}, due to the AC component of the capacitance $C_{c1} = C_c - C_{c0}$, and the current I_{cc0}, due to the capacitance C_{c0}. Then, introducing the notation

$$C_0 = C_1 + C_{c0},\quad \omega_0^2 = \frac{1}{L_1C_0},\quad d_0 = \omega_0 L_1 g_1,\quad \tau = \omega_0 t \tag{7}$$

and unifying Eqs. (1) and (7), we obtain a system of nonlinear equations for the behavior of the self-oscillator:

$$\begin{gathered} \ddot{U} + U = -d_0\dot{U} + \omega_0 L_1\frac{di_c}{d\tau} - \omega_0 M\frac{di_e}{d\tau} - \omega_0 L_1\frac{di_{cc1}}{d\tau}, \\ i_e = i_0\{(e^{aV_e} - 1) - \alpha_0(e^{aV_c} - 1)\}\left(1 - \frac{w_1}{w_0}\right), \\ i_c = i_0\{\alpha_0(e^{aV_c} - 1) - (e^{aV_c} - 1)\}\left(1 - \frac{w_1}{w_0}\right), \end{gathered} \tag{8}$$

$$i_{c1} = \omega_0 C_{co}\left[\left(1-\frac{U}{E_c'}\right)^{-1/2}-1\right]\dot{U},$$

$$V_e = k_1 U - i_{co} r_b'' + V_{bi} - \omega_0 L_2 (1-K^2)\frac{di_e}{d\tau} - r_b (i_e - i_c),$$

$$\dot{V}_{bi} + \frac{1}{\omega_0 \tau_{bi}}(V_{bi} - E_e') = -\frac{R}{\omega_0 \tau_{bi}} i_e.$$

$$i_{cc} = \omega_0 C_{co}\left(1-\frac{U}{E_c'}\right)^{-1/2}\dot{U}$$

$$w_1 = -d_{co}\left(1-\frac{U}{E_c'}\right)^{1/2}$$

It is impossible to reduce the system (8) to an equation with one unknown, U. We shall henceforth investigate particular cases, which can be analyzed by the methods of the theory of oscillations.

2. ANALYSIS OF SELF-OSCILLATOR IN THE CASE WHEN THE VOLTAGE INDUCED AT THE INPUT IS TOTALLY APPLIED TO THE EMITTER JUNCTION

Let the emf developed by the current i_e across inductance L_2 be much smaller than the junction voltage; the bias voltage is produced by a source having an infinitely small internal resistance, and the blocking capacitance C is sufficiently large. Neglecting at this stage the voltage $U''r_b$, the emitter-junction voltage can be written in the form

$$V_e = E_e + k_1 U. \tag{9}$$

Let us consider first the case when $C_0 \gg C_{c1}$, when it is possible to estimate most simply the influence of the base-width modulation, which determines the nonlinearity in the collector-junction resistance in the active region of the characteristics.

The self-oscillator equation will be written, according to (8) and (9), in the form:

$$\ddot{U} + U = -d_0\dot{U} + \xi_1\left\{k_1 e^{ak_1U}(\mu + \nu\sqrt{1-\beta}) - (e^{ak_1U} - e^{-aE_e})\frac{\zeta}{\sqrt{1-\beta}}\right\}\dot{U} - \\ - \xi_2 e^{-aE_{ce}}\left\{e^{aU}(\mu + \nu\sqrt{1-\beta}) - (e^{aU} - e^{aE_c})\frac{\zeta}{\sqrt{1-\beta}}\right\}\dot{U}, \tag{10}$$

where we use the notation

$$\xi_1 = \omega_0(\alpha_0 L_1 - M)\, aI_{e0};\ \ \xi_2 = \omega_0(L_1 - \alpha_0 M)\, aI_{e0};\ \ I_{e0} = i_0 e^{aE_e};$$
$$\nu = \frac{d_{c0}}{w_0};\ \ \mu = 1 - \nu;\ \ \beta = \frac{U}{E_c'};\ \ E_{ce} = E_c + E_e;\ \ \zeta = \frac{\nu}{2aE_c'}$$

We note that in Eq. (10) the terms containing ν and ζ take into account the effect of the base-width modulation.

To solve the problem we use the method of N.N. Bogolyubov. Putting

$$U = A\cos(\omega t + \varphi) = A\cos\theta, \tag{11}$$

we obtain the following symbolic equations for the amplitude and phase

$$\frac{d\theta}{d\tau} = 1, \tag{12}$$

$$\frac{dA}{d\tau} = -\frac{d_0 A}{2} + \frac{\xi_1 A}{2}\{k_1[\mu \ln t_2(\gamma) + \nu \ln t_3(\gamma)] - \zeta[\ln t_4(\gamma) - \\ - e^{aE_e}\ln t_5(\beta_0)]\} - \frac{\xi_2 A}{2} e^{-aE_{ce}}\{\mu \ln t_2(aA) + \nu \ln t_3(aA) - \\ - \zeta[\ln t_4(aA) - e^{aE_c}\ln t_5(\beta_0)]\}, \tag{13}$$

where

$$
\begin{gathered}
\ln t_2(\gamma) = \frac{2I_1(\gamma)}{\gamma}; \quad \ln t_3(\gamma) = \frac{2I_1(\gamma)}{\gamma} B_1(\beta_0) + \frac{4I_2(\gamma)}{\gamma} B_2(\beta_0) + \ldots ; \\
\ln t_4(\gamma) = \frac{2I_1(\gamma)}{\gamma} B_1'(\beta_0) + \frac{4I_2(\gamma)}{\gamma} B_2'(\beta_0) + \ldots ; \quad \ln t_5(\beta_0) = 1 + \frac{3}{32}\beta_0^2 + \ldots ; \\
B_1(\beta_0) = 1 - \frac{1}{32}\beta_0^2 - \ldots ; \quad B_2(\beta_0) = -\frac{\beta_0}{2}\left(1 + \frac{1}{16}\beta_0^2 - \ldots\right) \ldots ; \\
B_1'(\beta_0) = 1 + \frac{3}{64}\beta_0^2 - \ldots ; \quad B_2'(\beta_0) = \frac{\beta_0}{2}\left(1 + \frac{5}{16}\beta_0^2 + \ldots\right) \ldots ; \\
\gamma = ak_1A; \quad \beta_0 = A/E_c'.
\end{gathered} \tag{14}
$$

Here $I_1(\gamma)$, $I_2(\gamma)$ etc. are modified Bessel functions of the first kind. The limiting values of the functions B_1, B_2, ...; B'_1, B'_2, ...; $\ln t_5$ for $\beta_0 = 1$ are:

$$
\begin{gathered}
B_1 = 0.96; \quad B_2 = -0.55; \quad B_3 = -0.14; \ldots ; \\
B_1' = 1.2; \quad B_2' = 0.62; \quad B_3' = 0.46; \ldots ; \quad \ln t_5 = 1.2.
\end{gathered} \tag{15}
$$

The minimum value of k_1 is determined by the self-excitation condition

$$
\varepsilon = -d_0 + \xi_1 [k_1 - \zeta(1 - e^{-aE_e})] - \xi_2 [1 - \zeta(1 - e^{aE_c})] e^{-aE_{ce}} > 0.
$$

At room temperature, the parameters contained here have the following values [5, 6]:

$$
\begin{gathered}
a \simeq 40\mathbf{v}^{-1}; \quad \nu \simeq (1 - 2) \cdot 10^{-1}; \quad E_e \simeq 0.05 - 0.3\mathbf{v}; \\
E_c = 5 - 10\mathbf{v}; \quad \zeta \simeq 2.5 \cdot 10^{-4}; \quad e^{-aE_e} \leqslant 10^{-1} - 10^{-3}; \quad e^{aE_c} \simeq 10^{35} - 10^{160}
\end{gathered} \tag{16}
$$

Taking (16) into account the self-excitation condition can be represented in the form

$$
\varepsilon \simeq -d_0 + \xi_1 (k_1 - \zeta) > 0 \tag{17}
$$

or

$$
k_1 > \frac{g_{c0} + g_1}{g_{e0}}; \quad g_{c0} = \frac{I_{e0}\nu}{2E_{ce}}; \quad g_{e0} = aI_{e0}. \tag{18}
$$

Here g_{e0} and g_{c0} are the differential admittance of the emitter and collector at the working point.

Taking (15), (16), and (18) into account, we note that at small amplitudes the terms in the second curly bracket of (13) are negligibly small compared with the terms in the first, and at amplitudes comparable with E_c, one must take only the first term into account.

Examining the expression in the first curly bracket, we note that if g_{c0} is greater than or comparable with g_1, and if k_1 is close to ζ then the effect of the base-width modulation can lead to a limitation on the amplitude of the self-oscillations. Such a state can occur in the case of very weak regeneration.

In practice we usually do not use weak degrees of regeneration, for in this case the amplitude stability is low and the power delivered to the load is small. Therefore $k_1 \gg \zeta$ ($k_1 \geq 10^{-3}$) practically every time and we need leave in the curly bracket only the first term. This indicates that at sufficiently large couplings at low frequencies, the base-width modulation hardly influences the steady-state amplitude. We shall henceforth make use of this circumstance.

Thus, putting $\alpha_0 \simeq 1$ and $\zeta \ll k_1 \ll 1$, we obtain

$$
\frac{dA}{a\tau} = \frac{A}{2}\left\{-d_0 + (\varepsilon + d_0)\left[\frac{2I_1(\gamma)}{\gamma} - e^{-aE_{ce}} \frac{2I_1(aA)}{\gamma}\right]\right\} = F(A). \tag{19}
$$

In our case the amplitude limitation due to the input nonlinearity is impossible, since the condition for stability of the oscillations is found to be violated at any amplitude ($2I_2(\gamma) / / \gamma \geq 0$). Consequently, the limitation sets in as the result of cutoff of the collector voltage (the last term of the first half of (19) comes into play), and the form of the signal across the resonant circuit becomes different from harmonic (Fig. 2a), and the amplitude of the stable self-oscillations cannot be less than E_{ce}. This indicates a mode of hard excitation.

The amplitude variation up to the occurrence of cutoff is in accordance with the law

$$A = A_0 e^{\frac{\text{[illegible]}}{2}}, \tag{20}$$

where A_0 is the initial amplitude. As can be seen from Fig. 2a, the agreement between the experimental and calculated curves is fully satisfactory.

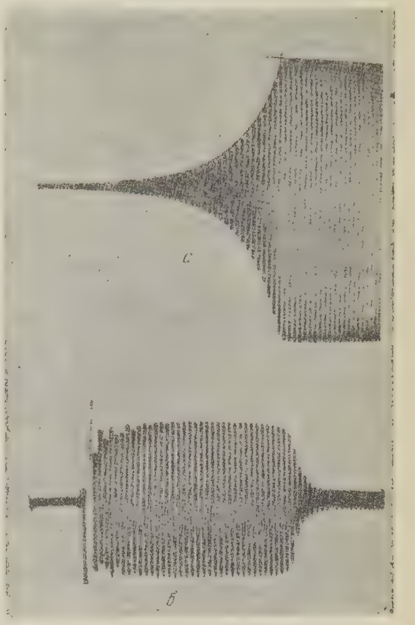


Fig. 2. a — Establishment of amplitude of self-oscillations (the calculation time of establishment is 38 cycles, while the experimental time is 36 cycles); b — establishment of frequency of self-oscillations (feedback-coil reactance much less than the emitter resistance ($\omega L_2 \simeq 1.8$ ohms; $R_{e0} \simeq 20$ ohms.

Experiment shows that in our case the amplitude of the steady-state oscillations cannot be considerably greater than E_{ce}. It is, therefore, natural to expect that magnitude of the optimal load at which maximum power transfer takes place is close to the breakdown load, as is indeed confirmed by the experimentally plotted dependence of the power on the load (Fig. 3). Assuming that the smallest amplitude of the stable oscillations is E_{ce}, we can obtain the breakdown load g_{br}. Recognizing that $e^{-aE_{ce}} I_1 (aE_{ce}) \ll I_1 (ak_1 E_{ce})$, we obtain from (19)

$$g_{br} \simeq \frac{2I_{eh}}{E_{ce}} I_1 (ak_1 E_{ce}) - g_1. \tag{21}$$

Fig. 4 shows the experimental and calculated variations of the breakdown resistance ($R_{br} = 1/g_{br}$) on the degree of coupling k_1.

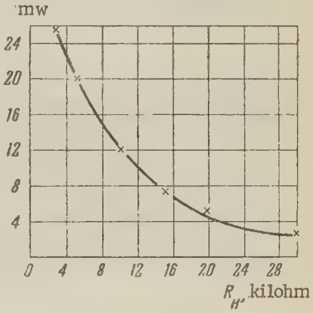


Fig. 3. Dependence of power delivered to load on the load resistance.

The first-approximation solution yields a phase which is constant during the buildup time of the self-oscillations (see (12)). However, an experimental investigation of the buildup of the frequency by an interference method indicates that the frequency is considerably reduced at the instant of appearance of cutoff (Fig. 2b).* This is attributed to the sharp increase in the contributions made by the harmonics to the signal spectrum.

The second approximation found by the method of N.N. Bogolyubov yields the following law for the establishment of the phase:

---

*The amplitude of the signal at the output of a phased interferometer connected after an amplitude limiter, at small frequency changes, has a value

$$e = A\Delta\omega(t)\tau_d \cos\omega(t)t,$$

where $\Delta\omega$ is the increment in frequency during the establishment of the oscillations; A — amplitude of the signal at the output of the limiter; τ_d — delay time of the signal in the second channel of the interferometer. Consequently, a constant frequency ($\Delta\omega = 0$ or const) in Fig. 2b corresponds to a signal of constant amplitude.

$$\frac{d\theta}{d\tau}=\left\{1-\frac{F(A)}{2A}\frac{dF(A)}{dA}-\left(\frac{d+d_0}{ak_1A}\right)^2\sum_{n=2}^{\infty}[I_n(aA)e^{-aE_{ce}}-I_n(ak_1A)]^2\right\}. \tag{22}$$

The third term (sum) in the right half of (22) determines the drop in oscillation frequency as a result of a reduction in the contribution of the harmonics, while the second term determines in practice the change in frequency with increasing amplitude prior to appearance of the cutoff. At the instant of cutoff F'(A) determines the stability of the amplitude), which should cause an increase in frequency, but a sharp increase in the weight of the harmonics nevertheless leads to a reduction in frequency.

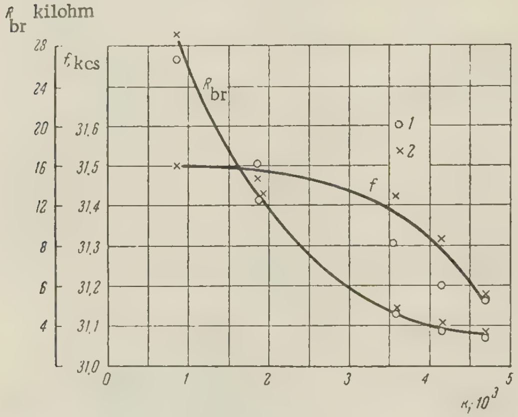


Fig. 4. Dependence of load and breakdown of oscillations on the degree of coupling

(1 — experimental point; 2 — calculated points) ($\omega L_2 \simeq 0.51$ ohms; $R_{e0} \simeq 25$ ohms; $L_1 = 23$ millihenry, $f_0 = 31.5$ kcs; transistor P16A; $F_{cr} = 1840$ kcs; $r_b = 40$ ohms; $\alpha_0 = 0.98$; $E_{ce} = 10$ v).

In the steady-state mode $E(A_{st}) = 0$. Therefore, the frequency of the steady-state self-oscillations is

$$\omega=\omega_0\left\{1-\left(\frac{\varepsilon+d_0}{ak_1A_{st}}\right)^2\sum_{n=2}^{\infty}[I_n(aA_{st})e^{-aE_{ce}}-I_n(ak_1A_{st})]^2\right\}. \tag{23}$$

The variation of frequency with degree of coupling, including the experimental points, is shown in Fig. 4 for optimum load.

EFFECT OF THE BASE RESISTANCE r''_b AND THE CURRENT i_{cc}

Let us evaluate the effect of the voltage developing across the base resistance r''_b by the current i_{cc}. The voltage on the emitter junction will be written in the form

$$V_e=E_e-k_1U-r''_bC_c\omega_0\dot{U}=E+v_e. \tag{24}$$

In addition, we retain in the first equation of the system (8) the term that takes into account the current in the capacitance C_{c1}.

Let us consider the case $av_e \leq 1$. We expand the function $\exp(aV_e)$ contained in the expression for the currents i_b and i_c in powers of av_e and confine ourselves to the first four terms. The resultant error is not more than 2%.

The solution of the problem by the method of N. N. Bogolyubov yields in the first approximation the following equations for the establishment of the amplitude and the phase

$$\begin{gathered}\frac{dA}{d\tau}=\frac{A}{2}\left\{-d_0+(\varepsilon+d_0)\left[G(\beta_0)-e^{-aE_{ce}}\frac{2I_1(aA)}{\gamma}\right]\right\},\\ \frac{d\theta}{d\tau}=1-\frac{\varepsilon+d_0}{2}K(\beta_0)\operatorname{tg}\psi_0-\frac{\eta}{2}\Phi(\beta_0),\end{gathered} \tag{25}$$

where

$$G=1+\frac{1}{8}\gamma_0^2\beta_0^2[1+\operatorname{tg}^2\psi_0\Phi_1(\beta_0)];$$

$$K=\Phi_2(\beta_0)+\gamma_0\beta_0^2\Phi_3(\beta_0)+\gamma_0^2\beta_0^2\left[\Phi_4(\beta_0)+\frac{1}{3}\operatorname{tg}^2\psi_0\Phi_5(\beta_0)\right];$$

$$\Phi_1 = 1 + \frac{1}{2}\beta_0^2 + \frac{5}{16}\beta_0^4 + \ldots; \quad \Phi_2 = 1 + \frac{3}{32}\beta_0^2 + \frac{3}{1024}\beta_0^4 + \ldots;$$

$$\Phi_3 = \frac{1}{8}\left[1 + \frac{5}{16}\beta_0^2 + \frac{325}{2048}\beta_0^4 + \ldots\right]; \quad \Phi_4 = \frac{1}{8}\left[1 + \frac{9}{16}\beta_0^2 + \frac{175}{2048}\beta_0^4 + \ldots\right];$$

$$\Phi_5 = \frac{3}{8}\left[1 + \frac{5}{16}\beta_0^2 + \frac{63}{1024}\beta_0^4 + \ldots\right]; \quad \Phi = \frac{3}{32}\beta_0^2 + \frac{35}{1024}\beta_0^4 + \ldots;$$

$$\gamma_0 = ak_1E_c'; \quad \eta = \frac{C_{c0}}{C_0}; \quad \operatorname{tg}\psi_0 = \frac{r_б'' C_{c0}\omega}{k_1}. \tag{26}$$

The limiting values of the functions Φ_1, Φ_2, . . . when $\beta_0 = 1$ are $\Phi_1 = 8$; $\Phi_2 = 1.2$; $\Phi_3 = 0.24$; $\Phi_4 = 0.24$; $\Phi_5 = 1.44$; $\Phi = 0.2$. The function $\eta/2\Phi(\beta_0)$ determines the change in frequency, due to the nonlinear capacitance C_{c1}.

It follows from the relations obtained, first, that the self-excitation condition remains the same as before. Second, the presence of r''_b leads to an increase in the rate of buildup of the amplitude during the process of establishment of the oscillations. Third, the voltage developed by the current i_{cc} across r''_b causes a reduction in the frequency of the self-oscillations with increasing amplitude. Fourth, the nonlinear capacitance C_{c1} also causes a reduction in the frequency during the process of establishment of the oscillations (when $C_1 = 0$ the reduction in the frequency is 20%).

All this is clear from physical considerations. At the initial instant the voltage across r''_b is much less than the voltage across the junction, and the self-excitation condition remains the same as before. With increasing oscillation amplitude, a feedback voltage is developed across r''_b that leads the resonant-circuit voltage by $\pi/2$. Thus, two voltages are effective in the input circuit of the transistor, shifted by $\pi/2$. The resultant junction emf, equal to the difference between k_1U and $U_{r''_b}$, is found to be shifted by an angle ψ relative to the resonant-circuit voltage. The collector current coincides in phase with the emf on the emitter junction and thus lags the voltage U. The latter circumstance reduces the frequency of the self-oscillations. Finally, an increase in the amplitude causes an increase in the first harmonic of the current through C_{c1}, which also leads to a decrease in the oscillation frequency.

We note that usually the product r''_bC_{c0} is a small quantity (less than 5×10^{-9} sec) and consequently at low frequencies, where $\tan\psi_0$ is close to zero, we can neglect the influence of r''_b on the operation of the self-oscillator.

CONCLUSION

A nonlinear analysis of the self-oscillations at low frequency produced with a junction transistor, leads to the following conclusions:

1. If the reactance of the feedback coils is less than the input resistance of the transistor, and if the coupling k_1 is limitingly small, the limitation of oscillations can be due to the base-modulation phenomenon. At values of k_1 used in practice, the limitation of amplitude is due to cutoff of the collector voltage, and the base modulation plays no role.
2. The presence of a base resistance leads to acceleration of the self-oscillation process and to a reduction in frequency with increasing amplitude.

The results of the theoretical investigations were confirmed by experiment.

The expressions obtained in this paper for the frequency, optimum load of the generator, and self-excitation conditions can be used in practice for design purposes.

REFERENCES

1. L.N. Kaptsov. Certain features of transistorized self-oscillators, Radiotekhnika i elektronika, 1957, 2, 9, 1127.
2. L.N. Kaptsov, K.Ya. Senatorov, O.I. Poplavaskaya. Concerning the establishment of oscillations in transistorized oscillators. Radiotekhnika i elektronika, 1959, 4, 9, 1557.
3. M.I. Kontorovich, S.G. Pellikan, V.N. Frolov. Investigation of a semiconductor self-oscillator, Scientific-Technical Bulletin No. 5, Radiophysics, Leningrad Polytechnic Institute, 1959.

4. N.N. Bogolyubov, Yu.A. Mitropol'skiy. Asymptotic methods in the theory of nonlinear oscillations, GIFML, 1958.
5. Yu.A. Kamenetskiy. Equivalent circuits of transistors. Collection of articles: Semiconductor Devices and Their Application, edited by Ya.A. Fedotov, No. II, 1957.
6. J.M. Early. Design Theory of Junction Transistors, Bell System Tech. J. 32, 271 (1953).
7. B.M. Vul. Dielectric properties of transition layers in semiconductors, ZhTF, Journal of Technical Physics, 1955, 25, 1, 3.
8. V.I. Samoilenko. Features of operation of semiconductor diodes and transistors at large voltages, Published by MAI, 1959.
9. W. Shockley. The theory of p-n junction semiconductors and p-n junction transistors, Bell System Techn. J., 1949, 28, 3, 435.
10. A.S. Maydanovskiy. Calculation of currents in an alloyed transistor for arbitrary amplitude of junction voltages. News of the Universities, Ministry of Higher Education USSR, Physics, 1961, 2, 34.

Received by editors 8 August 1960

INVESTIGATION OF NON-GAUSSIAN TRAJECTORIES IN THE FIELD OF AN ELECTROSTATIC QUADRUPOLE LENS*

A.M. Strashkevich

Equations are derived for the trajectories of a broad beam in the field of a quadrupole lens. It is shown that the analogous equations derived by Bernard and Hue are incorrect. Calculations are made for a specific system made up of four small spherical electrodes. It is shown that in a whole series of systems (both transgradient and congradient) the minimum deviation is experienced by trajectories which are symmetrical relative to the plane of symmetry of the field. This effect is negligibly small for the Gaussian region of the lens, but is considerable for the non-Gaussian one.

In an investigation of trajectories of charged particles beyond the limits of the Gaussian region in the fields of quadrupole electrostatic lenses, the main premises should be the equations of the trajectories of a broad beam with a straight-line axis in a field having two symmetry planes and two antisymmetry planes. In the nonrelativistic approximation, such equations were derived by Bernard and Hue [1]. However, as will be shown below, the equations they derived are incorrect.

To take into account the third-order terms in the trajector equations it is necessary to include in the expression for the potential in the field of a three-dimensional electrostatic quadrupole lens terms up to fourth order inclusive. Using the notation introduced in [2], we obtain

$$\Phi = \Phi_0 + C_z x^2 - C_z y^2 - \frac{1}{12} C_z'' x^4 + \frac{1}{12} C_z'' y^4, \tag{1}$$

---

*Reported at the Third All-Union Conference on Electron Microscopy, Leningrad, October, 1960.

where the prime denotes derivatives with respect to z. We take the axial trajectory of the beam to be a line coinciding with the z-axis. Then in the equations of the trajectories of a broad beam, obtained by the Greenberg method and given in [2] by Eqs. (16.15) and (16.26) we shall have (if we change the designations of the coordinate axes in accordance with the conditions formulated for our problem) $\lambda = 2C''_z$, $\mu = -2C_z$, $\nu = \epsilon = \sigma = \chi = \omega = \eta = 1/\rho = 1/\tau = 0$.

The equations assume in the nonrelativistic approximation the form

$$x'' - \frac{C_z}{\Phi_0}x = \frac{1}{\Phi_0}\Big(C_z xx'^2 + C_z xy'^2 - \frac{C_z^2}{\Phi_0}x^3 + + \frac{C_z^2}{\Phi_0}xy^2 - \frac{1}{2}C'_z x^2 x' + \frac{1}{2}C'_z x'y^2 - \frac{1}{6}C''_z x^3\Big), \tag{2}$$

$$y'' + \frac{C_z}{\Phi_0}y = -\frac{1}{\Phi_0}\Big(C_z yy'^2 + C_z x'^2 y + \frac{C_z^2}{\Phi_0}y^3 - - \frac{C_z^2}{\Phi_0}x^2 y - \frac{1}{2}C'_z y^2 y' + \frac{1}{2}C'_z x^2 y' - \frac{1}{6}C''_z y^3\Big). \tag{3}$$

One of these equations can be transformed into the other, if the sign of the function C_z is reversed and x and y interchanged (together with their derivatives). If we neglect in these equations the third-order terms, we obtain the paraxial equations, which are known in the literature. As is customary in the electron-optical theory of aberrations, we assume the solutions of the paraxial equations to be substituted in the right halves of (2) and (3) in lieu of x and y. Therefore, the second-order derivatives x'' and y'' in the right halves of the equations are replaced (respectively) by the expressions $\frac{C_z}{\Phi_0}x$, $-\frac{C_z}{\Phi_0}y$. Comparing (2) and (3) with analogous equations (4) of reference [1], we find that in (4) there are excessive terms containing $x'yy'$ and $xx'y'$. In the derivation described above, these terms drop out. The very same Eqs. (2) and (3) are obtained also directly from the ordinary formulas for the trajectory of a charged particle (see, for example, (4) in [3]:

$$2\Phi x'' = (1 + x'^2 + y'^2)\left(\frac{\partial\Phi}{\partial x} - x'\frac{\partial\Phi}{\partial z}\right), \tag{4}$$

$$2\Phi y'' = (1 + x'^2 + y'^2)\left(\frac{\partial\Phi}{\partial y} - y'\frac{\partial\Phi}{\partial z}\right). \tag{5}$$

In the approximation considered, which is necessary for the calculation of the third-order aberrations, the potential Φ, in the left halves of the equations, must be replaced by the expression $\Phi_0 + C_z x^2 - C_z y^2$; from the same expression we obtain $\partial\Phi/\partial z$ for the right halves of (4) and (5). On the other hand, the terms containing, in accordance with Eq. (1), the fourth powers of x and y are taken into account in the calculations of $\partial\Phi/\partial x$ and $\partial\Phi/\partial y$. The excess terms contained in the formulas of Bernard and Hue do not appear at all in the present derivation.

Nor are there any terms containing $x'yy'$ and $xx'y'$ in the more general equations for broad beams with straight-line axis in an arbitrary doubly symmetrical field. The derivation of the equations is quite analogous to the derivation of (2) and (3); these equations should, in particular, be used to calculate non-Gaussian trajectories for the quadrupole lenses having no antisymmetries. We have

$$x'' + \frac{\Phi'_z}{2\Phi_z}x' - \frac{V+W}{\Phi_z}x = \frac{1}{\Phi_z}\Big\{-\Big[\frac{(V+W)^2}{\Phi_z} + \frac{1}{6}(W''-2W_1)\Big]x^3 + + \frac{1}{2}\Big[\frac{(V+W)\Phi'_z}{\Phi_z} - (V'+W')\Big]x^2x' - \Big[\frac{V^2-W^2}{\Phi_z} + \frac{1}{2}(V''+2W_1)\Big]xy^2 + + \frac{1}{2}\Big[\frac{(V-W)\Phi'_z}{\Phi_z} - (V'-W'')\Big]x'y^2 + (V+W)xy'^2 + (V+W)xx'^2 - - \frac{1}{2}\Phi'_z x'^3 - \frac{1}{2}\Phi'_z x'y'^2\Big\} \tag{6}$$

(The notation follows Section II of [2]). The second equation of the trajectory is obtained if we replace all the x's in (6) by y's and vice versa, and at the same time we reverse the signs of the functions W and of their derivatives. In the case of axial field symmetry $W = 0$, $W_1 = \frac{3}{32} \Phi_z^{IV} = -\frac{3}{8} V''$. For the antisymmetrical quadrupole lens we have $\Phi_z = \Phi_0 = \text{const}$, $V = V' = V'' = W_1 = 0$, $W = C_z$, which converts the equations into form (2) and (3).

For a quadrupole lens formed by four small spheres (Fig. 1) the function C_z has the following form [4]:

$$C_z = 3Uba^2 (a^2 + z^2)^{-\frac{5}{2}}. \tag{7}$$

In this case the equations of the trajectory are best integrated by the method of power series. Let us consider a non-Gaussian trajectory in the field of our given system, the trajectory lying in the plane xz. We take as a unit scale the distance a. In the Gaussian approximation the equation for the trajectory will be

$$x'' + k\left(1 - \frac{5}{2} z^2 + \frac{35}{8} z^4 + \dots\right)x = 0, \tag{8}$$

where $k = -3 \frac{U}{\Phi_0} \frac{b}{a}$; in the calculations we assumed $k = 0.006$ and considered the interval $-0.5 \leq z \leq 0.5$. The function $(1 + z^2)^{-5/2}$ was expanded in a Maclaurin series so as to be able to investigate conveniently the dependence of the change of the deflection of the beam on the quantity $x_0 = (dx/dz)_{z=0}$. The point is that for many seemingly utterly unlike electron-optical systems (see below) one can prove one quite general statement: the angle of deflection is minimum in the case when the trajectory is symmetrical with respect to the symmetry plane of the field of the system. This rule is a generalization of the property of a refracting prism, known from optics. It is interesting to verify whether this takes place also for our case. The solution of (8) will be

$$\begin{aligned} x = x_0 (1 - 0.003000z^2 + 0.001250z^4 - 0.000876z^6 + \dots) + \\ + x_0' (z - 0.001000z^3 + 0.000750z^5 - 0.000625z^7 + \dots). \end{aligned} \tag{9}$$

We substitute (9) in the right half of (2) (we assume y to be equal to zero). Then (2) can be solved by varying the arbitrary constants. The integrals obtained in this manner are calculated in final form.

$$\Delta = \left(\frac{dx}{dz}\right)_{-0.5} - \left(\frac{dx}{dz}\right)_{0.5} = 0.005078x_0 + 0.03602\, x_0^3 + 0.1190\, x_0 x_0'^2. \tag{10}$$

From this expression (and from an analogous expression for a trajectory lying in the yz plane) it follows that a change in the sign of x'_0 (or y'_0) does not influence the angle of deviation which has, for fixed x_0 (or y_0), the smallest value when $x'_0 = 0$ (or $y'_0 = 0$). Fig. 2 shows curves for the dependence of Δ on x_0 at values x'_0, calculated by formula (10). Thus, this system has the property formulated above in the non-Gaussian region.

The same property is observed also for a deflecting homogeneous field; indeed, in the symbols explaining Fig. 3, we can write an equation for the trajectory that crosses the x-axis at the point x_0, in the form

$$x - x_0 = x_0' z + \frac{(1 + x_0'^2)\frac{d\Phi}{dx}}{4\Phi_0} z^2, \tag{11}$$

where Φ_0 is the point where the trajectory crosses the plane $z = 0$. Therefore

$$\left(\frac{dx}{dz}\right)_{z=-L} - \left(\frac{dx}{dz}\right)_{z=L} = -\frac{(1 + x_0'^2)\, L \frac{d\Phi}{dx}}{\Phi_0}. \tag{12}$$

We see from (12) that the angle between the initial and final directions of the beam is independent of the sign of x'_0 and has a minimum when $x'_0 = 0$, i.e., when the trajectory is

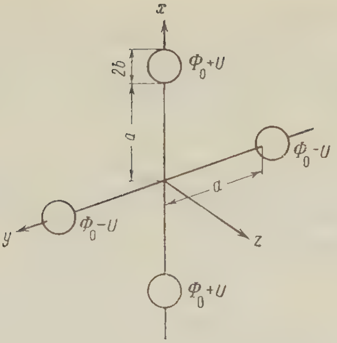


Fig. 1

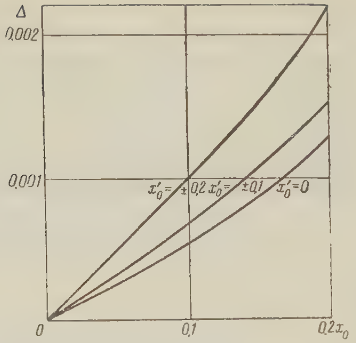


Fig. 2

symmetrical with respect to the plane z = 0. It is easy to show that this property is possessed by a homogeneous field also at relativistic particle velocities.

Both the quadrupole lens and a flat capacitor are transgradient systems with respect to the charged-particle motion considered above. However, the fact that the beam has a minimum deflection in the case of symmetry is not connected with this circumstance. Indeed, we can cite an example of a congradient deflection system ([5], pp. 45 and 142), which also has the same property. Fig. 4 shows the trajectories obtained by graphic-analytic means for the plane field of this system whose electrodes are shown by the heavy lines, plotted with an electrolytic trough. The dashed lines show the grid, and the potentials are given in arbitrary units, in view of the similarity law. The minimum change in the direction is possessed by the symmetrical trajectory too, which crosses the grid at a right angle: this trajectory is deflected by 80° whereas the neighboring nonsymmetrical trajectories are deflected by 90 and 106°.

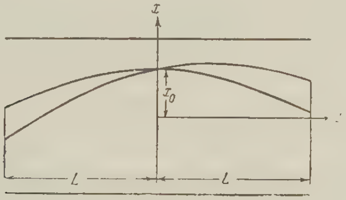


Fig. 3

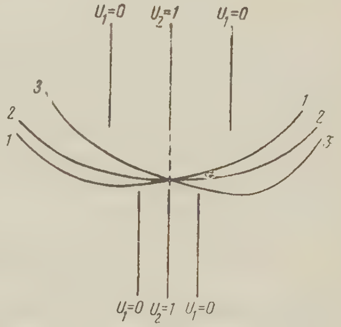


Fig. 4

In conclusion let us show that the effect investigated above does not take place in the Gaussian region for single lens of either quadrupole or other types. In fact, regarding the symmetry plane of an arbitrary single lens, we find that for any specified l (see Fig. 5) the angle of deflection is independent of the form of the trajectory:

$$\Delta = \alpha + \beta = l\left(\frac{1}{a_1} + \frac{1}{a_2}\right) = \text{const}, \tag{13}$$

since the brackets contain the value of the optical strength, which is constant for the lens. This is valid for both thin lens (when the principal planes H_1 and H_2 coincide) and thick lenses.

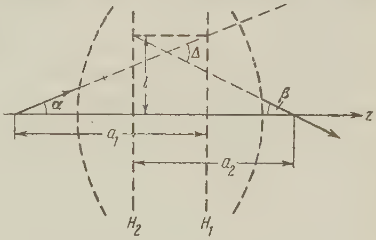


Fig. 5

REFERENCES

1. M.Y. Bernard, J. Hue. C.r. Acad. sci., 1956, 243, 23, 1852.
2. A.M. Strashkevich. Electron optics of electrostatic fields having no axial symmetry, GIFML, 1959.
3. A. Recknagel, Z. Phys., 1938, 111, 61.
4. G.D. Archard, Proc. Phys. Soc. B, 1955, 68, 11, 817.
5. E. Bruch and O. Scherzer. Geometrische Elektronenoptik, Springer, Berlin, 1934.

Received by editors 3 May 1961

PROCEDURE FOR ENERGY ANALYSIS OF THERMIONIC ELECTRONS UNDER PULSED EMISSION CONDITIONS

V.N. Lutskiy

An electric circuit is developed with which it is possible to measure the energy distribution functions of thermionic electrons in pulsed and "quasi-static" modes while plotting the volt-ampere characteristics, and which also permits an investigation of the deformation of the distribution function in time when operating in a pulsed mode.

An analyzer with sufficient resolution was developed. The procedure was verified with a tungsten cathode and both the circuit and the analyzer were found to have correct operation; a check made on a semiconductor cathode has shown that the expected effect, the investigation of which brought about the development of the procedure, exceeds by many times the accuracy of the method.

1. FORMULATION OF THE PROBLEM

One of the possible causes of the lack of saturation in the volt-ampere characteristics of semiconductor thermionic cathodes is the deformation of the equilibrium distribution function of the electrons over the energies in the body of the emitter, as occurs under the influence of the internal electric field produced in the near-surface layer as a result of the flow of the emission current. The possiblity of such "heating" of the electron gas and the form of the nonequilibrium distribution function were considered theoretically in [1] and [2]. The expression obtained by the authors of these papers for the distribution function was used in [3] to calculate the emission current as a function of the temperature and of the electric field.

The heating of the electron gas was observed experimentally by Dubretsov and Bulyginskiy [4]. The electron temperature was determined from the slope of the linear portion of the delay curves plotted on a semi-logarithmic scale. The measurements were carried out in a three-electrode cylindrical system (cathode — grid — collector) with oxide cathode and continuous operation at a constant positive grid potential. The measurement results have shown that the electron temperature is higher than the lattice temperature, and that the difference between these temperatures has a tendency to increase with increasing lattice temperature.

It is interesting to obtain a direct connection between the absence of saturation of the volt-ampere characteristics and the deformation of the distribution function, i.e., to measure the distribution of the thermionic electrons over the energies simultaneously with measuring the volt-ampere characteristics, and to obtain a more complete picture of the phenomenon it is desirable to turn to the region of highest anode voltages possible.

From the point of view of broadening the range of current densities and anode voltages, with account taken of the limited power dissipated by the anode, it is advantageous to carry out the measurements under pulsed conditions. In addition, the elucidation of the features of pulsed emission of semiconductor thermionic cathodes in the presence of strong electric fields on the cathode surface, as well as a comparative analysis of the results of the pulsed and continuous modes, are of independent interest. In particular, it would be of importance to the understanding of the phenomena leading to the drop in the pulsed emission current with time to investigate the distribution functions at different instants of time during the period of action of the pulse, with a wide variation of the pulse duration desirable. The first attempt in this direction is contained in the paper by Morgulis and Zingerman [5], but the analyzer used by the authors, which was suitable for the main purpose of their investigation, could not give the true form of the delay curves, owing to electron-optical distortion.

Thus, the problem is to develop a method with which, to measure the energy distribution function of the thermionic electrons, simultaneously with plotting the pulsed and static volt-ampere characteristics, in both the continuous mode and during various instants of time during the action of the pulse.

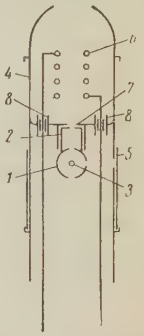


Fig. 1. System of electrodes of the analyzer.

2. CONSTRUCTION OF ANALYZER

To measure the energy distribution function of the thermionic electrons, we have employed the delay-curve method in a spherical-capacitor system. This method enables us to base the analysis on total energies; the resolution of the method is determined by the ratio of the radii of the internal and external electrodes and by the maximum angle of emission of the electrons, in accordance with formulas [6, 7]

$$\frac{\Delta v}{v} = \frac{r^2}{R^2}\sin^2\varphi, \tag{1}$$

where Δv — interval of unresolvable velocities in electron volts; v — total velocity of electron in electron volts; r — radius of internal electrode; R — radius of external electrode; φ — maximum angle of electron emission.

By virtue of the fact that in our case the system must be of the three-electrode type (cathode — anode — electrode), we must take v to mean the summary velocity, made up of the initial velocity and a velocity acquired in the transit space under the action of the anode voltage. At a high anode potential, reaching to 10 kv, this circumstance strongly increases the interval of unresolvable velocities, so that the usual 1% accuracy employed in such systems is found to be insufficient at an electrode radius ratio equal to 10.

It is possible to increase the resolving power of the instrument by reducing the maximum angle of emission of the electrons with sufficiently strong diaphragming and attenuation of the effect of the scattering lens, unavoidably occurring near the exit diaphragm of the anode. These considerations, together with considerations involving the choice of optimum geometry for the production of a maximum field at the cathode surface at a given anode potential, were indeed the basic premises for the choice of the velocity-analyzer construction.

The construction chosen is a modification of the system used by Muller and Young [8] to measure the distribution of field-emission electrons by energies. The general agreement the analyzer had is shown in Fig. 1. Here 1 — anode in the form of a cylinder 5 mm in diameter and 10 mm high, having an upper diaphragm of diameter 0.03 — 0.04 cm, holes 0.8 mm in diameter on the side for the measurement of the cathode temperature, and similar holes below for the centering of the filament (cathode); 2 — cylinder 3.5 mm in diameter and 3 — 4 mm high, in the bottom of which is contained a second diaphragm 0.03 — 0.04 cm in diameter; 3 — filament-like cathode; 4 — outside cylinder, terminated with a hemisphere 16 mm in diameter with a hole 8 mm in diameter; 5 — sliding cylinder, covering a slot 5 × 1 mm in the outer cylinder, intended for the measurement of the temperature; 6 — tungsten coil for outgassing the assembly (during the measurements the coil was at the same potential as the anode); 7 — screen protecting the working part of the analyzer against the penetration of scattered electrons. The leads of the coil are insulated from the screen with the aid of quartz tubes, covered, to prevent charging of the quartz, with tantalum cylinders 8 whose lengths were considerably greater than the lengths of the quartz insulators.

The entire assembly was made of tantalum. The collector was the spherical surface of a bulb on which a conducting layer of platinum was sputtered. The sputtering was such that upper and lower hemispheres were insulated from each other; in the lower hemisphere a window was left for optical pyrometry. The current of the scattered electrons on the lower hemisphere was completed outside the measuring circuit. The diameter of the collector was 190 mm, i.e., the ratio of the electrode radii was $R/r = 11.9$. The error due to the nonsphericity of the field in the construction chosen by us, as shown in references [9, 10], was insignificant. A hemisphere with an exit diaphragm of sufficiently large diameter, in conjunction with a paraxial electron beam of small diameter, ensured small scattering of the electrons and at the same time eliminated the scattering lens formed in the region of small anode diaphragms.

3. ELECTRIC CIRCUIT

The gist of the method by which it is possible to measure the energy distribution function of the thermionic electrons and the alteration of this function with time, simultaneously with plotting the volt-ampere characteristics, is clear from the description of the electric circuit used for the measurements.

a) Pulsed mode. In the measurement circuit described, an important factor is the use of high-voltage pulses with positive polarity, since the main part of the measuring circuit is concentrated in the collector-cathode circuit and must be grounded.

Usually the circuit with accumulating capacitor, used to produce high-voltage pulses, produces voltages of negative polarity. The use of a high-voltage pulse pulse transformer is excluded, in view of the wide range of variation of the pulse duration. The use of a circuit comprising a thyratron with a long line is complicated and inconvenient.

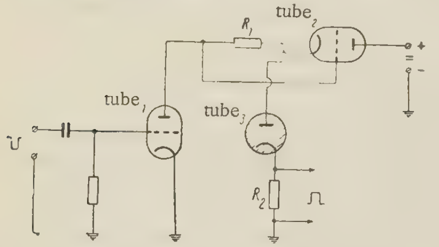


Fig. 2 Principal diagram of high-voltage pulse generator of high-voltage process with positive polarity.

We, therefore, chose a simple circuit (Fig. 2), which enabled us to obtain a positive pulse without a stringent limitation on the repetition frequency and on the pulse duration, and with practically full utilization of the power-supply voltage.

The circuit operates as follows. In the absence of a signal on the grid of tube 1, this tube conducts. The plate current of tube 1 produces on R$_1$ a voltage drop which almost cuts off tube 2, so that practically all the source voltage is applied to tube 2. The voltage applied between the cathode of tube 2 and the housing is insufficient to ignite the gas-discharge tube 3 and the voltage and the voltage drop on R$_2$ is zero. When a negative-polarity pulse with several hundred volts amplitude is applied to the grid of tube 1, tube 1 cuts off while tube 2 conducts, and after the firing of tube 3 (disregarding the small drop across tubes 2 and 3) the supply voltage is applied in this case across resistor R$_2$. After determination of the triggering pulse, the circuit returns to the initial state.

The circuit described produces at the output pulses of positive polarity, the duration of which vary from 50 to 1,000 microseconds, with a repetition frequency from 40 to 1,000 pulses per second, and the voltage is continuously adjustable from zero to 10 kv.

The measuring circuit is shown in Fig. 3.

For the measurements in the pulsed mode, switch S$_1$ is set in position I. A positive pulse is applied to the anode of the analyzer. The voltage pulse proportional to the cathode current is picked off the resistance group R$_{col}$ and is fed to the input of the vertical-deflection amplifier of the first beam of a double-beam oscilloscope. A pulse proportional to the anode voltage is applied to the input of the vertical amplifier of the second beam from divider D$_1$. A voltage proportional to the collector current pulse, the magnitude of which is determined by the density of the emission current, by the form of the distribution function of the emitted electrons, by the resolution of the instrument, and by the collector potential, is picked off the resistor group R$_{col}$ and is fed to the input of the vertical amplifier of the first oscilloscope beam. Connected to the input of the vertical amplifier of the second beam is a sawtooth voltage pulse, applied between the collector and the cathode, with the negative terminal connected to the collector. In this case the form of the current pulse on the collector if

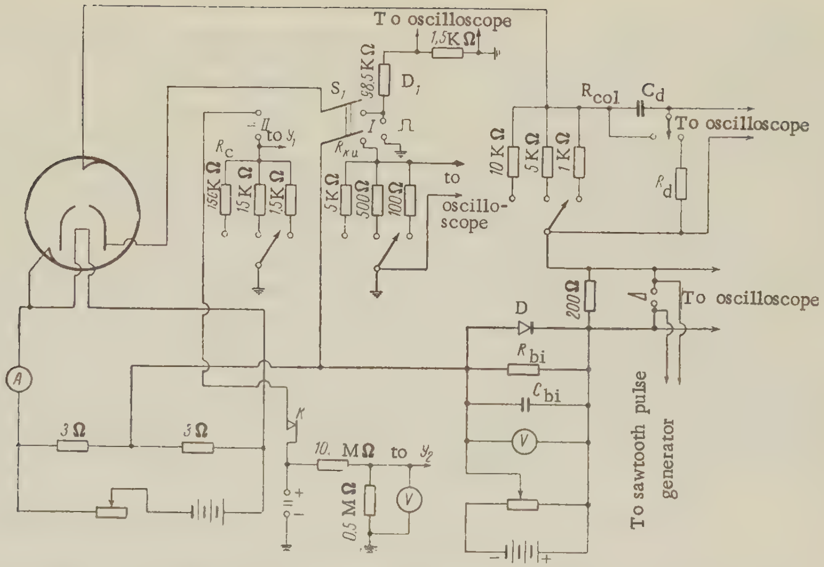


Fig. 3. Electric measurement circuit.

the sawtooth pulse has good linearity and if the oscilloscope sweep voltage has good linearity, taking into account the voltage drop across the resistances in the collector — cathode circuit, is the delay curve, i.e., the function $i_{col} = f(V_c)$ (V_c is the collector potential). Inasmuch as the oscillograph displays simultaneously with the delay curve also the voltage pulse applied to the collector, the requirement of linearity of V_c and linearity of the sweep voltage is not obligatory (the motion of both oscillograph beams is controlled by the same sweep voltage).

The duration of the sawtooth pulse is chosen to be 7.5 and 25 microseconds, i.e., a small fraction of the duration of the rectangular pulse of anode voltage. By applying a sawtooth voltage with different delay relative to the start of the rectangular pulse, it is possible to plot the delay curve at any instant of time of operation of the anode-voltage pulse.

The synchronization system was produced in this case in the following fashion: an external pulse of required frequency triggered a high-voltage square-wave generator and the delay block of a sawtooth pulse generator. The oscillograph sweep was triggered by the delay pulse simultaneously with application of the sawtooth voltage pulse.

Using a differentiating network consisting of a capacitance $C_d = 100$ micromicrofarad and a resistance $R_d = 500 - 1000$ ohms, one can obtain directly the energy distribution curve of the thermionic electrons. The quantities C_2 and R_d were determined by trial: if the differentiating network operates correctly, the sawtooth pulse applied to its output should be converted into a rectangular pulse, provided it has sufficiently good linearity. The use of electric differentiation, however, entails a considerable reduction in the sensitivity and can be used, therefore, only at sufficiently large emission-current densities.

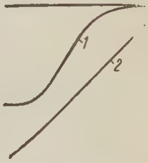


Fig. 4. Typical form of oscillograms: 1 — collector current (null line above); 2 — sawtooth collector voltage pulse.

To compensate for the contact difference of potentials and for the voltage drop due to the cathode and collector currents, a constant positive bias is applied to the collector, and its magnitude can be continuously varied from zero to 40 volts. The potentiometer from which the bias voltage is picked off was shunted by a capacitance $C_{bi} = 30$ micromicrofarads and a semiconductor diode D to ensure a low time constant in the capacitor-discharge circuit. The presence of these elements made the AC impedance of the bias circuit negligibly small.

The time constants, which are determined both by the input capacitance of the oscilloscope and by the internal capacitance of the analyzer, did not exceed several tenths of a microsecond; the frequency characteristics of the oscilloscope amplifier made it possible to obtain oscillograms without distortion of the true form of the investigated relationships.

A typical oscillogram is shown in Fig. 4.

b) Quasistatic mode. The customarily used static-measurement modes did not permit us to employ anode voltages greater than 1 — 2 kv, owing to the strong heating of the anode. We were able to eliminate this difficulty in the following fashion.

When the anode voltage is turned on, the heating time, depending on the current picked off and on the applied voltage, fluctuates approximately from 0.5 second to several seconds. With relation to the processes that occur in the cathode after the anode voltage is turned on, this time is much longer than the time required for the stationary state of the cathode to be established, and, consequently, measurements made during that interval can be regarded as being made with the cathode in continuous operation.

This measurement mode is realized in position II of switch S_1 (oil switch K, which closes the circuit during the required time interval. Part of the anode voltage is fed from the voltage divider directly to the vertical-deflection plates of the second beam of the oscilloscope. The voltage applied to the deflection plates of the first beam is proportional to the anode current, and is picked off the group of resistances R_c.

Measurements in the collector circuits are carried out in the same fashion as in the pulsed mode (for the sake of simplicity, Fig. 3 does not show the switching circuit).

By applying the anode voltage for a short interval of time, we were able to operate with voltages on the order of 5 — 6 kv, without overheating the cathode.

4. TEST OF THE PROCEDURE WITH TUNGSTEN CATHODES DISCUSSION OF RESULTS

The main purpose of the control measurements made with the tungsten cathode was a comparison of the experimental delay curves with the theoretical ones. If the volt-ampere characteristics are correctly measured, this will certainly reduce uncertainty considerably.

Distortion of the form of the delay curves can be due both to the quality of the analyzer and to incorrect operation of the electric circuitry. Distortion due to the operation of the circuit can immediately be eliminated from consideration, for, in addition to carefully checking the circuit elements with regards to the time constants and the frequency characteristics, we established that the form of the delay curves plotted in the ordinary manner (without the use of our circuit) always coincided with the results obtained with the circuit.

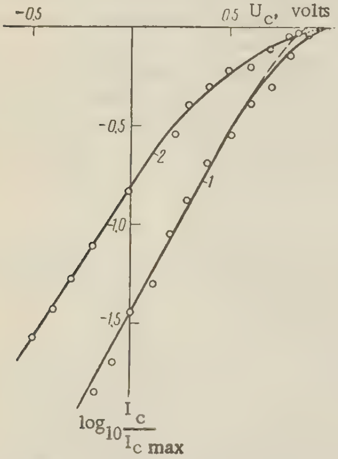


Fig. 5. Delay curves in semilogarithmic scale for tungsten cathode (static mode)

1 — T = 2,650° K, U_a = 1 kv; 2 — T = 3,180° K, U_a = 0.5 kev. The dashed curve is the theoretical relationship for T = 2,650°.

Thus, the correctness of the results was essentially determined by the quality of the analyzer. An idea of the quality of the analyzer can be gained by comparing the results of measurements made with a tungsten cathode whose thermionic-electron energy distribution function can be regarded as Maxwellian [11] with the theoretical curve for a given temperature.

Fig. 5 shows the delay curve plotted semilogarithmically, for tungsten at 2,650° K and at 1 kv on the anode (curve 1). The dashed curve is the theoretical one. We see from Fig. 5 that the discrepancy occurs at low electron energies. This discrepance is not exclusively the consequence of lack of sufficient resolution, but can also be due to secondary emission from the collector. The presence of secondary emission is confirmed by the form of the characteristic in the region past saturation (see Fig. 6). Fig. 6 shows on the linear scale the same curve as in the preceding case. The correction for the secondary emission can be made, for example, by the method described in [8].

A direct estimate of the resolution from the comparison of the experimental curve with the theoretical one at low electron energies is, therefore, difficult. It is more correct to compare the theoretical and experimental characteristics, using a temperature determined

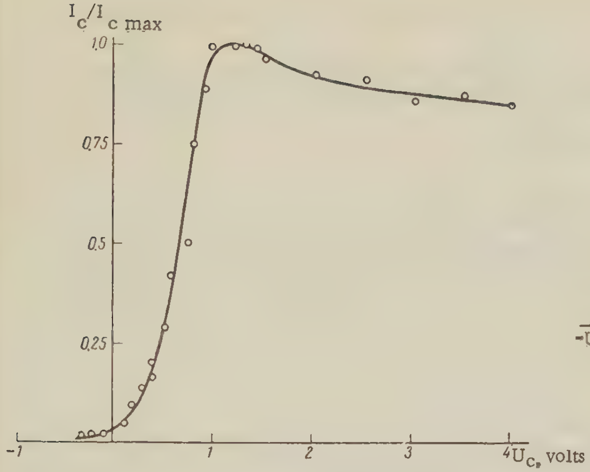


Fig. 6. Delay curve for tungsten cathode in static mode at U_a = 1 kv, T = 2,650° K.

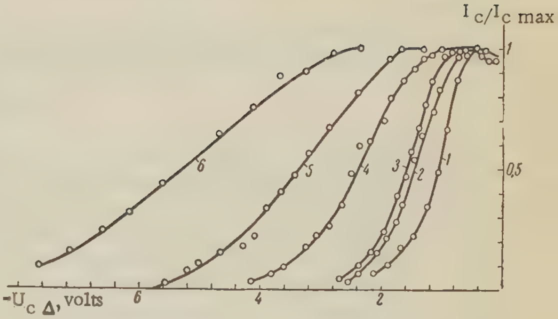


Fig. 7. Delay curves in pulse mode for different anode voltages (curve 1 pertains to tungsten, all others to a semi-conductor cathode): 1 — U_a = 9 kv; 2 — U_a = 0.5 kv; 3 — U_a = 1 kv; 4 — U_a = 4 kv; 5 — U_a = 7 kv; 6 — U_a = 9 kv.

from the slope of the straight-line portion of the delay curve on the semilogarithmic plot. In the case of good resolution the determination of temperature from the slope of the experimental delay curve yields sensible values. If the resolution is somewhat deteriorated, say as a result of deformations in the wiring occurring during the outgassing, the calculated values of the temperature are greatly increased and reach more than 5,000° K. This yields a sensitive criterion of the quality of the analyzer. Although this criterion does not yield the exact value of the interval of unresolved velocities, it does guarantee smallness of this interval, and as a consequence, correctness of the determination of the investigated distribution function.

The same remark can be made concerning the width of the delay curve as the current is changed, say, by two orders of magnitude. In many cases the difference between the theoretical and measured widths of the delay curve was less than 0.05 volts, and the deterioration in resolution due to increasing the anode diaphragm diameter to 1 mm led to an increase in width by 2 — 3 times.

Fig. 5 shows, for the sake of comparison, the delay curve for T = 3,180° K at U_a = = 500 v. In both cases the temperature calculated from the delay curves was compared with the temperature determined from the tables on the basis of the emission-current density, with allowance for the smallness of the emitting surface of the cathode, owing to the large temperature gradient along the cathode.

The fact that the theoretical width of the delay curve agrees well with the measured one, and the sensible temperature values obtained from the slope of the delay curve are also evidence, on top of all the above, that the use of non-pulsed heating of the cathode, and also the potential drop along the filament, do not exert an appreciable influence on the quality of the analyzer. The point is apparently that when such strong electric fields are produced at the cathode surface, the effect of the magnetic field due to the filament current is small and is practically uniform for any part of the spectrum, so that the distribution is "transferred" to the anode-diaphragm region without distortion. As regards the potential drop along the filament over a length equal to the diameter of the diaphragm, it does not exceed 0.01 — 0.02 volts, and naturally does not exert any influence on the resolution of the analyzer.

We also checked the procedure using several types of semiconductor cathodes. Fig. 7 shows results obtained in the pulsed mode. The abscissas represent the ac component of the collector potential.

The constant positive collector potential, not shown in the figure, increased with increasing anode voltage and reached 24 v at 9 kv. Curve 1 pertains to tungsten at 2500° K and U_a = 9 kv. Other curves, plotted for different anode voltages, pertain to a semiconductor cathode at 1500°. It is seen that with increasing anode voltage the distribution function is greatly changed. Curve 1 shows that this change cannot be attributed to the properties of the analyzer or of the measuring circuit. The effect is reduced with increasing temperature.

An analysis of the results obtained with a semi-conductor cathode is beyond the scope of our paper.

In conclusion, I am deeply grateful to B.M. Tsarev for general guidance of the work and to M.I. Yelinson for numerous useful discussions and systematic help with the work.

REFERENCES

1. B.I. Davydov. ZhETF, 1936, 6, 463; 1937, 7, 1069.
2. B.I. Davydov and I.M. Shmushkevich. ZhETF, 1940, 10, 1043.
3. S.M. Levitin, ZhETF. (Journal of Technical Physics), 1953, 23, 1700.
4. L.N. Dobretsov and D.G. Bulyginskiy. ZhTF, 1946, 26, 5, 977.
5. N.D. Morgulis and Ya.P. Zingerman. Transactions of Conference on Cathode Electronics, Kiev, 1952, p. 69.
6. P.I. Lukirskii. On the photoeffect, GTTI, 1933.
7. P.I. Lukirskii. Uspekhi, Fiz, Nauk, 1945, 27, 2, 199.
8. R. Young, and E. Muller. Phys. Rev., 1959, 113, 1.
9. N.A. Soboleva. Radiotekhnika i elektronika, 1958, 3, 3, 339.
10. D.A. Ganichev and K.G. Utkin. Paper delivered to the Interdepartmental Seminar on Cathode Electronics, Eleventh Session, 1 December 1958.
11. L.N. Dobretsov and P. Ya. Uvarov, ZhTF, 1952, 22, 175.

Received by editors 21 December 1960

BRIEF COMMUNICATIONS

ESTIMATE OF FREQUENCY OF NARROW-BAND SIGNAL AGAINST A WHITE-NOISE BACKGROUND

V.I. Bunimovich, V.A. Morozov

1. The purpose of the present communication is to show that the problem of estimating the frequency of a signal received against the background of white noise is equivalent to the problem of determining the position of the maximum of a function representing the superposition of a stationary Gaussian noise and a certain regular process (see Section 6 below).*

2. We assume that the expression for the signal is

$$S(t) = C(t)\cos[\omega t - \theta(t)] \tag{1}$$

and assume further, as usual, that the envelope and the "phase" $\theta(t)$ of the signal are slowly varying functions of the time.

We carry out the estimate by the maximum-likelihood method [1]. Here, as is well known, we can simultaneously assume that the estimate is obtained by the method of maximum of a posteriori probability [2, 3], if the form of the curve of the a priori distribution of the parameter is sufficiently close to rectangular. One must naturally bear in mind the different meanings of the estimates obtained by these methods.

3. On the basis of the well known expression for the conditional probability density of an oscillation obtained against the background of white noise (in the case of a signal "known accurately," see [4]), we can write for the likelihood function of the parameter

$$L(\omega) = \exp\left[-\frac{1}{N_0}\int_0^T (x - S)^2\, dt\right], \tag{2}$$

where T is the period of observation; N_0 is the spectral density of the noise; x is the incoming oscillation (sampling function).

An idea concerning the parameter ω, on which the signal S depends, is obtained from the "observed" function x.

The expression in the exponents of the right half of (2) is represented in the form of a sum of three terms. The only term dependent on the parameter of interest to us is $\frac{2}{N_0}\int_0^T xS\, dt$. In lieu of the latter it is convenient to examine the proportional quantity

$$u = \frac{q}{E}\int_0^T xS\, dt, \tag{3}$$

---

*In the case of a signal with "unknown" phase, the problem of estimating the frequency, and also of estimating the delay, reduces to the determination of the distribution of the position of the maxima of the envelope of the same function.

where E is the signal energy; q characterizes the signal-to-noise ratio:

$$E = \int_0^T S^2\,dt;\quad q = \sqrt{\frac{2E}{N_0}}. \tag{4}$$

It is clear that the likelihood function (2) assumes a maximum value simultaneously with expression (3).

4. We can write for the functions x and S

$$\begin{gathered} x = n\,(t) + C\,(t)\cos\,[\omega_0 t - \theta\,(t)], \\ S = C\,(t)\cos\,[\omega t] - \theta\,(t) = C(t)\cos\left[\frac{\nu t}{T} + \omega_0 t - \theta\,(t)\right], \end{gathered} \tag{5}$$

where n(t) is the white noise and ω is the estimated parameter. We take as an estimate the value of the parameter at which the right half of (3) is a maximum (we have in mind in general, naturally, the greatest maximum). The value of ν in (5) is

$$\nu = (\omega - \omega_0)\,T. \tag{6}$$

This quantity, proportional to the deviation of ω from the true value of the frequency ωt, will henceforth be taken to be the estimated parameter.

5. We can now write, obviously,

$$u = u_N\,(\nu) + u_S\,(\nu), \tag{7}$$

where

$$u_N\,(\nu) = \frac{q}{E}\int_0^T n\,(t)\,C\,(t)\cos\left[\frac{\nu t}{T} + \omega_0 t - \theta\,(t)\right]dt;$$

$$\begin{aligned} u_S(\nu) &= \frac{q}{E}\int_0^T C^2\,(t)\cos\,[\omega_0 t - \theta\,(t)]\cos\left[\frac{\nu t}{T} + \omega_0 t - \theta\,(t)\right]dt = \\ &= \frac{q}{2E}\int_0^T C^2\,(t)\cos\frac{\nu t}{T}\,dt + \frac{q}{2E}\int_0^T C^2\,(t)\cos\left[\left(2\omega_0 + \frac{\nu}{T}\right)t - 2\theta\,(t)\right]dt \end{aligned} \tag{8}$$

or approximately

$$u_S\,(\nu) = \frac{q}{2E}\int_0^T C^2\,(t)\cos\frac{\nu t}{T}\,dt. \tag{9}$$

The component $u_S(\nu)$ is a regular function of the parameter ν. On the other hand, the component $u_N(\nu)$ is a random function of ν (we note that the representation (7) corresponds to an expansion into a "signal" and noise function, as was done in [2].

On the basis of the well known expression

$$\overline{n\,(t)\,n\,(t')} = \frac{N_0}{2}\delta\,(t' - t) \tag{10}$$

for the correlation function of a white noise, we obtain for the correlation function of the random "process" $u_N(\nu)$ (see (8)) the following expression (see also (4)):

$$\begin{aligned} \overline{u_N\,(\nu)\,u_N\,(\nu + \vartheta)} &= \frac{1}{E}\int_0^T C^2\,(t)\cos\left[\frac{\nu t}{T} + \omega_0 t - \theta\,(t)\right]\cos\left[(\nu + \vartheta)\frac{t}{T} + \omega_0 t - \theta\,(t)\right]dt = \\ &= \frac{1}{2E}\int_0^T C^2\,(t)\cos\frac{\vartheta t}{T}\,dt + \frac{1}{2E}\int_0^T C^2\,(t)\cos\left[\left(2\omega_0 + \frac{2\nu}{T} + \frac{\vartheta}{T}\right)t - 2\theta\,(t)\right]dt \end{aligned}$$

or, approximately

$$\overline{u_N(\nu)\,u_N(\nu+\vartheta)} = \overline{u_N u_{N\vartheta}} = \frac{1}{2E}\int_0^T C^2(t)\cos\frac{\vartheta t}{T}\,dT, \tag{11}$$

hence

$$\overline{u^2} = 1, \tag{11a}$$

since

$$E = \int_0^T S^2(t)\,dt \simeq \frac{1}{2}\int_0^T C^2(t)\,dt \text{ (см. (4) и (1))}.$$

Thus, the random component of the function u (see (7)) is a stationary Gaussian random function of the parameter, with a mean value equal to zero, with a dispersion equal to unity, and with a correlation coefficient defined by the form of the envelope of the signal in accordance with formula (11). In lieu of (11) and (9) we can write

$$\overline{u_N u_{N\vartheta}} = r(\vartheta),\quad u_S(\nu) = qr(\nu) \text{where} \ r(x) = \frac{1}{2E}\int_0^T C^2(t)\cos\frac{xt}{T}\,dt. \tag{12}$$

In the particular case when C = const we have

$$u_S(\nu) = q\frac{\sin\nu}{\nu} \text{ ("signal")}$$

(12a)

$$\overline{u_N u_{N\vartheta}} = \frac{\sin\vartheta}{\vartheta} \text{ (correlation coefficient of "noise")}$$

We note incidentally that it is easy to obtain from (11) (or (12)) the following expression for the spectral density F(p) of the random process $u_N(\nu)$:

$$F(p) = \frac{T}{2E}C^2(Tp). \tag{13}$$

6. The foregoing arguments show that the problem of estimating the frequency of a narrowband signal (with a single unknown parameter) against the background of white noise is equivalent to the problem of the distribution of the position of the greatest maximum of a function representing the superposition of a Gaussian stationary random process and a regular signal. The correlation function of the random process and the function representing the signal have one and the same form (see (12)) and are determined by the form of the envelope of the actual signal (1).

To be sure, the problem of the distribution of the position of the maxima has not yet been solved in general. Such a representation, however, in addition to its clarity (compared with the initial formulation) is also useful because it leads to a simulation method that yields experimentally an estimate of the frequency. In our case it is easy to obtain an asymptotic expression for the estimate (one valid for large values of q, see [2] and Section 7 below), but even for this case it is of undoubted interest to determine experimentally the applicability region of the asymptotic expression (these experiments are not cited in the present communication).

7. Let us touch upon now, for the sake of completeness, the question of the asymptotic estimate of the frequency, an estimate which is valid for large values of the signal/noise ratio, or more accurately for sufficiently large values of the quantity q (see (4)). In this case the greatest maximum of the function $u(\nu)$ with probability close to unity will be the maximum which is closest to the origin $\nu = 0$, i.e., the estimate will be the smallest value of ν satisfying the equation

$$u(\nu) = u_N + qr(\nu) = \max \tag{14}$$

[see (7) and (12)], i.e., the smallest root of the equation

$$u'(\nu) = u'_N(\nu) + qr'(\nu) = 0. \tag{15}$$

It would be natural to attempt to derive an expression for the smallest root of Eq. (15) at large values of q in the form of an expansion in powers of the small parameter $\epsilon = 1/q$:

$$\nu = \nu(\varepsilon) = \nu(0) + \nu'(0)\,\varepsilon + \ldots\,. \tag{16}$$

It is easy to see [see (15) and (12)] that $\nu(0) = 0$; by differentiating (15) with respect to ϵ we readily obtain

$$\nu'(0) = \frac{u'_N(0)}{-r''(0)},$$

where [see (12)]

$$-r''(0) = \overline{u'^2_N} = \frac{1}{2T^2E}\int_0^T t^2C^2(t)\,dt = \frac{\overline{t^2}}{T^2}, \tag{17}$$

$\overline{t^2}$ is defined by the well-known expression

$$\overline{t^2} = \frac{\int_0^T t^2C^2(t)\,dt}{\int_0^T C^2(t)\,dt}. \tag{18}$$

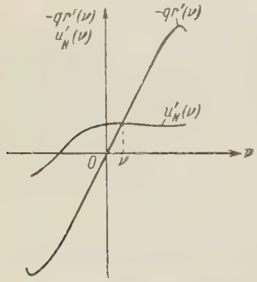
The first approximation for the estimate of the parameter ν [see (16)] will be, on the basis of the foregoing,

$$\nu = \nu'(0)\,\varepsilon = \frac{u'_N(0)}{-qr''(0)} = \frac{T^2u'_N(0)}{\overline{t^2}q} \tag{19}$$

and in the particular case C = const we have $\overline{t^2} = T^2/3$ and

$$\nu = \frac{3u'_N(0)}{q}. \tag{19a}$$

Noting further that by virtue of (12) we have $\overline{u'^2_N} = -r''(0)$, we obtain on the basis of (19) and (17) the following expression for the dispersion of the estimate



$$\overline{\nu^2} = \frac{T^2}{\overline{t^2}q^2} \tag{20}$$

and when C = const we get

$$\overline{\nu^2} = \frac{3}{q^2} = \frac{3N_0}{C^2T}, \quad \overline{(\omega - \omega_0)^2} = \frac{3N_0}{C^2T^3}. \tag{20a}$$

We note that the estimates (20) and (20a) are unbiased and asymptotically best [1]. The expression (20a) agrees with that obtained in reference [2].

We note that by successive differentiation of both halves of (15) with respect to ϵ [see (16)] we can readily obtain further approximations for the estimate, shown in the form of expansion (16), but in practice this is meaningless, since the probability of the expansion (16) being meaningful [i.e., the probability of such realizations of the function $x = x(5)$, at which the expansion (16) is valid], will in general be close to unity for large q, as is shown in (19).

8. Without discussing further possible methods of deriving (19) and its refinements, let us indicate briefly, in conclusion, a geometrical interpretation of this formula. This interpretation clarifies, so to speak, the meaning of the approximation obtained and can also be considered as an elementary method of derivation.

The figure shows schematically the plots of the functions $-qr'(\nu)$ and $u'_N(\nu)$. The abscissa of the point of intersection of these curves determines the sought root of Eq. (15). It is clear from the figure that at sufficiently large values of q, we have

$$u'_N(0) \simeq u'_N(\nu) \simeq \nu \, (-qr'(\nu))'_{\nu=0} = -\nu q r''(0).$$

The last equation is equivalent to (19).

REFERENCES

1. H. Cramer. Mathematical Methods of Statistics, Princeton University Press, 1956.
2. V.A. Kotel'nikov. The Theory of Potential Interference Immunity, Dissertation, MEI, 1946; GEI, 1956.
3. P.M. Woodward, Probability and Information Theory with Applications to Radar, McGraw-Hill, New York, 1953.
4. W. Peterson, T.G. Birdsall, and W.C. Fox. The theory of signal detectability, IRE Trans., 1954, PGIT-4, 171.

Received by editors 10 May 1961

CALCULATION OF RELIABILITY OF TROPOSPHERIC COMMUNICATIONS SYSTEM IN THE PRESENCE OF CORRELATED FADING

A.V. Prosin

In UHF communication using long-distance tropospheric propagation, various types of separated reception are used to overcome the rapid fading of the signals. The reliability of the communication system with respect to the rapid fading in the case of multiple reception will be characterized by the percentage of time during which at least one of the amplitudes of the separated signals exceeds the corresponding threshold reception level

$$H_n = (1 - S_n)\,100\%, \tag{1}$$

where S_n is the probability of having all the amplitudes of the separated signals below the threshold reception levels u_{t1}, u_{t2}, . . . , u_{tn}:

$$S_n = \underbrace{\int_0^{u_{t1}} \int_0^{u_{t2}} \cdots \int_0^{u_{tn}}}_{n} W_n(u_1, u_2, \ldots, u_n)\, du_1, du_2 \ldots, du_n. \tag{2}$$

In Eq. (2), $W_n(u_1, \ldots, u_n)$ is the n-dimensional differential distribution function of the instantaneous values of the amplitudes of the separated signals at the receiver inputs.

Expressions (1) and (2) determine the reliability of the communication system in the case of separated reception with switch over, when the stronger signal is chosen. Inasmuch as the gain obtained by using the method of signal addition over the method of switching is known, expression (1) also enables us to determine the reliability of the communication system in which signal addition is used.

The purpose of the present communication is to determine the reliability of the long-distance communication system in the case of duplicate signal reception. It is assumed in the calcuation that the field at the reception point is made up of a fixed and scattered components, and that the fadings in the different separation channels are correlated.

In accordance with (1) and (2), to calculate the reliability of duplicate reception it is sufficient to determine the two-dimensional integral distribution function

$$S_2 = \int_0^{u_{t1}} \int_0^{u_{t2}} W_2 (u_1, u_2)\, du_1 du_2, \tag{3}$$

where the differential distribution function (in the presence of a constant and random components of the signal at the reception point) is

$$W_2 (u_1, u_2) = \frac{u_1 u_2}{\sigma^4 (1 - R^2)} e^{-\frac{u_1^2 + u_2^2 + u_{01}^2 + u_{02}^2 - 2u_{01} u_{02} R}{2\sigma_2 (1 - R^2)}} \times$$
$$\times \sum_{m=0}^{\infty} \varepsilon_m I_m \left[\frac{R u_1 u_2}{\sigma^2 (1 - R^2)} \right] I_m \left[\frac{u_{01} - R u_{02}}{\sigma^2 (1 - R^2)} u_1 \right] I_m \left[\frac{u_{02} - R u_{01}}{\sigma^2 (1 - R^2)} u_2 \right]. \tag{4}$$

In formula (4), u_1 and u_2 or u_{01} and u_{02} are respectively the amplitudes of the summary and of the constant signals, acting in the first and second separation channels, respectively; σ^2 — dispersion of the orthogonal components of the random vectors of the signals: $R^2 = R_{13}^2 + R_{14}^2$; R_{13} and R_{14} are the coefficients of correlation of the orthogonal components of the random of the field; $\epsilon_0 = 1$; $\epsilon_m = 2$ for $m > 0$; I_m — is the Bessel function of order m of pure imaginary argument.

We shall henceforth confine ourselves to an analysis of the case when $u_{01} = u_{02} = u_0$, which is of practical importance. By substituting (4) in (3) we obtain a double integral, which is not expressed in terms of elementary functions. Let us find the approximate solution of (3). We represent the Bessel functions of pure imaginary argument in (4) as a series

$$I_n (x) = \sum_{k=0}^{\infty} \frac{1}{k!\, \Gamma (n + k + 1)} \left(\frac{x}{2} \right)^{n+2k}, \tag{5}$$

where Γ is the gamma function

When the Bessel functions are replaced by the series (5), the variables u_1 and u_2 in (3) separate, and consequently it is sufficient to calculate, in lieu of the double integral, the single integrals that result from the separation of the variables. After integrating (3) we obtain

$$S_2 = (1 - R^2)\, e^{-\frac{2z}{1-R}} \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{\varepsilon_m R^{m+2k}}{\Gamma (k + 1)\, \Gamma (m + k + 1)} z^m \times$$
$$\times \left[\sum_{l=0}^{\infty} \frac{\Gamma (m + k + l + 1)}{\Gamma (l + 1)\, \Gamma (m + l + 1)} z^l I (u_1, p_1) \right] \left[\sum_{n=0}^{\infty} \frac{\Gamma (m + k + n + 1)}{\Gamma (n + 1)\, \Gamma (m + n + 1)} z^n I (u_2, p_2) \right]. \tag{6}$$

Where $I (u, p)$ is the incomplete function [1]:

$$I (u, p) = \frac{1}{\Gamma (p + 1)} \int_0^{u\sqrt{p+1}} x^p e^{-x}\, dx; \tag{7}$$

$$u_1 = \frac{u_{t_1}^2}{2\sigma^2(1-R^2)\sqrt{m+k+l+1}}; \tag{8}$$

$$p_1 = m+k+l; \tag{9}$$

$$u_2 = \frac{u_{t_2}^2}{2\sigma^2(1-R^2)\sqrt{m+k+n+1}}; \tag{10}$$

$$p_2 = m+k+n; \tag{11}$$

$$z = \gamma^2\frac{1-R}{1+R}; \tag{12}$$

The ratio of the constant and random signal powers is

$$\gamma^2 = \frac{u_0^2}{2\sigma^2}. \tag{13}$$

From (7) we have when the receiver input is only the random field component

$$S_2 = (1-R^2)\sum_{k=0}^{\infty} R^{2k}\, I\,(u_1, p_1)\, I\,(u_2, p_2). \tag{14}$$

When the receivers have identical threshold properties $(u_{t_1} = u_{t_2} = u_t)$, we obtain from (6) and (14)

$$S_2 = (1-R^2)\, e^{-\frac{2z}{1-R}} \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \frac{\varepsilon_m R^{m+2k}}{\Gamma(k+1)\,\Gamma(m+k+1)}\, z^m \times$$

$$\times \left[\sum_{l=0}^{\infty} \frac{\Gamma(m+n+l+1)}{\Gamma(l+1)\,\Gamma(m+l+1)}\, z^l\, I\,(u, p)\right]^2, \tag{15}$$

$$S_2 = (1-R^2)\sum_{k=0}^{\infty} [R^k I\,(u, p)]^2. \tag{16}$$

The tables of the functions (14) and (16) are given in [2]. Let us consider some particular cases that follow from Eq. (15).

1. As R → 0

$$S_2 = S_1^2 = \left[e^{-\gamma^2}\sum_{l=0}^{\infty}\frac{1}{\Gamma(l+1)}\gamma^{2l} I\,(u, p)\right]^2. \tag{17}$$

2. As R → 0 and $\gamma^2 \to 0$

$$S_2 = S_1^2 = [I\,(u, p)]^2 = \left(1 - e^{-\frac{u_t^2}{2\sigma^2}}\right)^2. \tag{18}$$

3. As R → 1

$$S_2 = S_1 = e^{-\gamma^2}\sum_{l=0}^{\infty}\frac{1}{\Gamma(l+1)}\gamma^{2l} I\,(u, p). \tag{19}$$

4. As R → 1 and $\gamma^2 \to 0$

$$S_2 = S_1 = I\,(u, p) = 1 - e^{-\frac{u_t^2}{2\sigma^2}} \tag{20}$$

Expressions (19) and (20) are one-dimensional distribution functions of the signal amplitudes for the generalized Rayleigh law and for the Rayleigh law proper, respectively. If it is necessary to plot the dependence $S_2 = f(S_1)$ for $\gamma = 0$, we must put in (14) or (16)

$$u = -\frac{\ln(1 - S_1)}{(1 - R^2)\sqrt{k+1}}. \tag{21}$$

5. If the following inequalities are satisfied

$$\frac{R^2 u_t^4}{4\sigma^4 (1 - R^2)^2} \ll 1, \quad \frac{u_0^2 u_t^2}{4\sigma^2 (1 + R)^2} \ll 1 \tag{22}$$

we have

$$S_2 = \left\{ e^{-\frac{2z}{1-R}} \left[1 - e^{-\frac{1}{1-R^2} \frac{u_t^2}{2\sigma^2}} \right] \right\}^2. \tag{23}$$

In the presence of a random field component at the reception point we have

$$S_2 = \left[1 - e^{-\frac{1}{1-R^2} \frac{u_t^2}{2\sigma^2}} \right]^2. \tag{24}$$

It is interesting to note that formulas (23) and (24) have outwardly the same form as formula (18), calculated for the case $R = 0$ and $\gamma^2 = 0$. The correlation of the signals and of the DC component of the field is accounted for in (23) and (24) by the factors $[-2z(1-R)^{-1}]$ and $(1-R^2)^{-1}$.

In long-distance communication we frequently have $\gamma^2 < 1$. From (1) and (16) it follows that if a random field exists at the reception point, the reliability of the system with duplicate reception depends appreciably on the correlation coefficient between the separated signals only when $R > 0.6$. Consequently, in practice we very frequently have $z \ll 1$. For $z \ll 1$ we obtain from (6)

$$S_2 = e^{-\frac{2z}{1-R}} \left\{ (1 - R^2) \sum_{k=0}^{\infty} R^{2k} I(u_1, p_1) I(u_2, p_2) \right\}. \tag{25}$$

When $u_{t1} = u_{t2} = u_t$ we have

$$S_2 = e^{\frac{2z}{1-R}} \left\{ (1 - R^2) \sum_{k=0}^{\infty} [R^k I(u, p)]^2 \right\}. \tag{26}$$

The functions in the curly brackets of (25) and (26) coincide with the functions (14) and (16), which are tabulated in [2].

An analysis of (23) — (26) shows that the reliability of a communication system in the case of separated reception increases appreciably as the power of the constant signal increases compared with the power of the random signal. The reliability is least when only the random component of the field is present at the point of reception. An increase in the coefficient of correlation between the separated signals greatly reduces the reliability of the communication.

REFERENCES

1. E.E. Slutskii. Tables for the calculation of the incomplete Γ function and the probability functions χ^2, Edited by A.N. Kolmogorov, Academy of Sciences Press, 1950.
2. N.A. Huttly. Use of the tetrachoric cross-correlation in hypotheses concerning autocorrelated fading signals, statistical methods in radio wave propagation, Proc. of a symposium held at the University of California, Los Angeles, June 18-20, 1958, p. 154-175.

3. Staras. Diversity reception with correlated signals, J. Appl. Phys., 1957, 27, 3.

Received by editors 27 March 1961

CASCADED BRIDGES

A. L. Fel'dshteyn

In [1] we investigated optimal cascaded directional couplers, made up of coupled lines (Fig. 1). Let us examine the conditions under which these systems turn into bridges.

A bridge is defined as a directional coupler in which the voltages in the arms are equal in absolute value and have a constant phase shift.

The cascaded couplers considered in [1] do not produce a constant phase shift between arms 2 and 3 of a four-port network, and consequently are not bridges in the sense indicated above, even when the power is equally divided.

We indicate below the conditions under which a cascaded directional coupler becomes a bridge, and tabulate certain typical problems involving the calculation of three-decibel bridges of this type.

A cascaded directional coupler has, for any number of steps, a theoretically perfect directivity and is matched at all frequencies [1]. To define it completely it is, therefore, sufficient to specify the modulus and the phase of the coefficient of voltage division between arms 2 and 3:

$$T_{41} = M = \frac{S_{12}}{S_{13}}, \tag{1}$$

where S_{12} and S_{13} are elements of the scattering matrix.

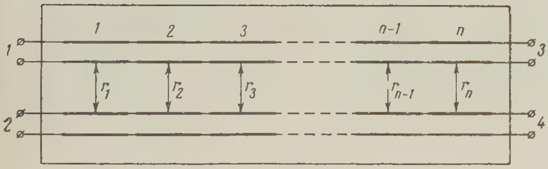


Fig. 1

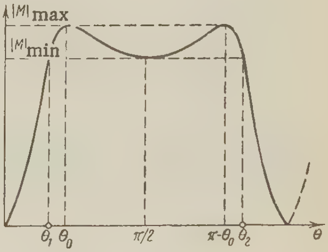


Fig. 2

We shall stipulate that

$$|M| = 1, \qquad \arg M = \frac{\pi}{2} \tag{2}$$

at all frequencies, and starting from the wave transfer matrix of one step

$$[T] = \begin{bmatrix} \cos\theta + j\rho\sin\theta & 0 & 0 & -jr\sin\theta \\ 0 & \cos\theta + j\rho\sin\theta & -jr\sin\theta & 0 \\ 0 & jr\sin\theta & \cos\theta - j\rho\sin\theta & 0 \\ jr\sin\theta & 0 & 0 & \cos\theta - j\rho\sin\theta \end{bmatrix} \tag{3}$$

(ρ and r are respectively the wave resistance and the coupling resistance, normalized to the load resistance R) we investigate the product of several matrices such as (3). By induction we reach the conclusion that a cascaded directional coupler becomes a bridge when the number of cascades is odd and the system is completely symmetrical.

Equations for the calculation of a three-step bridge with equal power division have the form

$$\varepsilon^3 - \frac{27}{4}\varepsilon + \frac{27}{4}|M|_{min} = 0 \tag{4}$$

and

$$(\varepsilon - 2r_1)(1 + 2r_1^2) - 2r_1\sqrt{1 + r_1^2}\sqrt{1 + (\varepsilon - 2r_1)^2} = |M|_{min} \tag{5}$$

where

$$\varepsilon = 2r_1 + r_2, \tag{6}$$

and the meaning of $|M|_{min}$ is clear from Fig. 2.

Eq. (4) can be solved analytically, and (5) graphically. Substituting the obtained value of r_1 in (6), we obtain r_2. The numerical results of the calculation are summarized in the table.

Nominal transfer attenuation of the bridge, db	Maximum deviation from nominal value of δ (tolerance), db	Normalized coupling resistances of the cascades		Coefficient of band overlap λ_2/λ_1
		$r_1 = r_3$	r_2	
3	0.1	0.136	1.401	2.450
	0.2	0.154	1.437	3.00
	0.5	0.192	1.492	4.064
	1.0	0.236	1.530	5.320

The amplitude-frequency characteristics of a three-cascade bridge (with tolerance δ = 0.5 db) is shown in Fig. 3 (solid curve). The dashed curve is the characteristic of a two-step Chebyshev-type directional coupler, calculated in accordance with [1] for the same value of the tolerance. The two characteristics are practically the same. Thus, to attain identical results with respect to matching, directivity, and power division it is advantageous to use two cascades in a coupler and three in a bridge. This can be readily understood if it is recognized that the bridge must satisfy the requirement of having not only constant power division with respect to the amplitude-frequency characteristic, but also constancy with respect to the phase-frequency characteristic. In a cascaded directional coupler, on the other hand, argM depends on the frequency (Fig. 4).

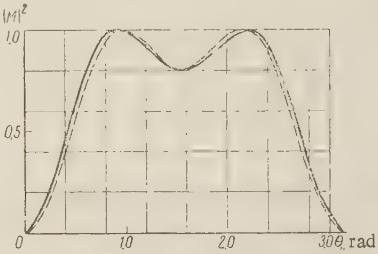


Fig. 3

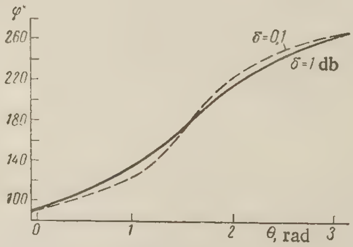


Fig. 4

It is expedient to construct the bridge or directional coupler in the form of a strip [2] or coaxial [3] coupled line; the length of each cascade is $l = \lambda_{av}/4$, where λ_{av} is the wavelength corresponding to the central frequency of the band.

The author is grateful to O.I. Mazepova and E.V. Solov'eva for help in the numerical calculations.

REFERENCES

1. A.L. Fel'dshteyn. Synthesis of cascaded directional couplers, Radiotekhnika i elektronika, 1961, 6, 2, 234.
2. J.G.K. Shimizy. Tripled directional couplers, IRE Wescon Convention Record, 1957, August, part I, 4.
3. A.L. Fel'dshteyn. Coupled inhomogeneous lines, Radiotekhnika, 1961, 16, 5, 7.

Submitted to editors 11 October 1960

VARIATIONAL PRINCIPLE FOR HOLLOW SYSTEMS WITH ANISOTROPIC MEDIUM

V.V. Nikol'skiy

The variational principle can serve as the basis for the construction of various methods of calculating irregular gyrotropic resonators, waveguides, and waveguide transformers, as well as other hollow systems with inhomogeneous isotropic filling. In particular, it can be verified that the matrix formulations arising in eigenfunction expansion of the fields [4, 5] (or their like) are obtained by employing the Ritz method in the corresponding variational problem.

Convenient variational relations for resonators and waveguides, as is well known, were published by Berk [1]. However, Berk's formulas are suitable only when the dielectric-constant and permeability tensors of the internal medium are Hermitian. We obtain below similar expressions, which are free of the foregoing limitations and at the same time are much simpler than the Hauser functionals [2, 3].

Let us take a hollow electromagnetic resonator with a perfectly conducting shell S, delineating a volume V, filled with an inhomogeneous anisotropic medium. The permeability and dielectric constant of the internal medium (tensors, functions of the coordinates) are denoted by the symbols μ and ϵ.

Writing down the Maxwell equations

$$\begin{aligned} \operatorname{rot} \vec{H} &= j\omega\varepsilon\vec{E}, \\ \operatorname{rot} \vec{E} &= -j\omega\mu\vec{H}, \end{aligned} \tag{1}$$

we start out from the assumption that the variational principle for the natural frequency ω can be formulated also when μ and ϵ are non-Hermitian [i.e., when energy is either absorbed or generated (regenerated), as in parametric systems]. Then, as follows from the general premises of the variational method (see, for example [6], Chapter IX), one must introduce the "conjugate" quantities $\tilde{\mathrm{E}}$ and $\tilde{\mathrm{H}}$ — the solutions of equations which can, in particular, be obtained by verification of the variational principle. Multiplying both lines of (1) by $\tilde{\vec{\mathrm{E}}}$ and $\tilde{\vec{\mathrm{H}}}$ respectively, and integrating over V, we obtain

$$\omega = j\,\frac{\int\limits_V \left(\tilde{\vec{H}}\,\mathrm{rot}\,\vec{E} - \tilde{\vec{E}}\,\mathrm{rot}\,\vec{H}\right) dv}{\int\limits_V \left(\tilde{\vec{H}}\mu\vec{H} + \tilde{\vec{E}}\varepsilon\,\vec{E}\right) dv}, \tag{2}$$

or abbreviated

$$\omega = \frac{M}{N}. \tag{2a}$$

Regarding (2) as a functional, we set up the variation

$$\delta\,(M - \omega N) \equiv \Delta. \tag{3}$$

We see that

$$\Delta = j\int\limits_V \left(\delta\tilde{\vec{H}}\,\mathrm{rot}\,\vec{E} + \tilde{\vec{H}}\,\mathrm{rot}\,\delta\vec{E} - \delta\tilde{\vec{E}}\,\mathrm{rot}\,\vec{H} - \tilde{\vec{E}}\,\mathrm{rot}\,\delta\vec{H}\right) dv -$$

$$-\,\omega\int\limits_V \left(\delta\tilde{\vec{H}}\mu\vec{H} + \tilde{\vec{H}}\mu\delta\vec{H} + \delta\tilde{\vec{E}}\varepsilon\vec{E} + \tilde{\vec{E}}\varepsilon\delta\vec{E}\right) dv$$

or

$$\Delta = j\int\limits_V \delta\,\tilde{\vec{H}}\left(\mathrm{rot}\,\vec{E} + j\omega\mu\vec{H}\right) dv - j\int\limits_V \delta\tilde{\vec{E}}\left(\mathrm{rot}\,\vec{H} - j\omega\varepsilon\vec{E}\right) dv +$$

$$+\,j\int\limits_V \delta\vec{E}\left(\mathrm{rot}\,\tilde{\vec{H}} + j\omega\tilde{\vec{E}}\varepsilon\right) dv - j\int\limits_V \delta\vec{H}\left(\mathrm{rot}\,\tilde{\vec{E}} - j\omega\tilde{\vec{H}}\mu\right) dv + j\int\limits_S \left\{[\delta\vec{E}, \tilde{\vec{H}}] + [\tilde{\vec{E}}, \delta\vec{H}]\right\}\vec{ds}. \tag{4}$$

This result shows that $\triangle$ vanishes under the following conditions
1) The fields $\vec{E}$ and $\vec{H}$ are solutions of Maxwell's equation (1).
2) The variation $\delta\,\vec{E}$ obeys the ordinary homogeneous boundary condition

$$[\delta\vec{E}, \vec{n}_0] = 0 \text{ for } S. \tag{5}$$

3) The fields $\tilde{\vec{E}}$ and $\tilde{\vec{H}}$ are solutions of the equations

$$\begin{aligned} \mathrm{rot}\tilde{\vec{H}} &= -\,j\omega\tilde{\vec{E}}\varepsilon, \\ \mathrm{rot}\,\tilde{\vec{E}} &= j\omega\tilde{\vec{H}}\mu, \end{aligned} \tag{6}$$

which can be called the conjugate Maxwell equations (these were the equations used by Hauser in [2, 3]), under the same usual condition

$$[\tilde{\vec{E}}, \vec{n}_0] = 0 \text{ for } S. \tag{7}$$

The variational principle has thus been established, and in the case of Hermitian μ and ϵ the formula (2) becomes the well-known formula given by Berk [1], for then

$$\tilde{\vec{H}} = \vec{H} \text{ and } \tilde{\vec{E}} = \vec{E}^*.$$

It is likewise easy to check that the variational principle for a cavity with non-Hermitian permeability and permittivity, is also expressed by the relations

$$\omega^2 = \frac{\int\limits_V \mathrm{rot}\,\tilde{\vec{E}}\,\frac{1}{\mu}\,\mathrm{rot}\,\vec{E}\,dv}{\int\limits_V \tilde{\vec{E}}\varepsilon\vec{E}dv}, \tag{8}$$

and

$$\omega^2 = \frac{\int\limits_V \operatorname{rot} \widetilde{\vec{H}} \frac{1}{\varepsilon} \operatorname{rot} \vec{H}\, dv}{\int\limits_V \widetilde{\vec{H}} \mu \vec{H} dv}, \tag{9}$$

which in turn go in the Hermitian case into the formulas of [1]. At the same time, functionals which are written somewhat differently may prove to be more convenient.* For example, in lieu of (8) we write

$$\omega^2 = \frac{\int\limits_V \widetilde{\vec{E}} \frac{1}{\varepsilon} \operatorname{rot} \frac{1}{\mu} \operatorname{rot} \vec{E}\, dv}{\int\limits_V \widetilde{\vec{E}} \vec{E} dv} \tag{8a}$$

$$\omega^2 = \frac{\int\limits_V \widetilde{\vec{D}} \operatorname{rot} \frac{1}{\mu} \operatorname{rot} \frac{1}{\varepsilon} \vec{D}\, dv}{\int\limits_V \widetilde{\vec{D}} \vec{D} dv} \tag{8b}$$

(we do not write out the analogous formulas with magnetic vectors).

We can obtain in the same manner several variational relationships for a waveguide. We confine ourselves to a generalization of the most interesting formula of reference [1] to the case of non-Hermitian constants. Namely, introducing the previous conjugate fields $\widetilde{\vec{E}}$ and $\widetilde{\vec{H}}$ (solutions of (6) under boundary condition (7) on the shell of the waveguide), we can readily obtain from Maxwell's equation the following functional for the propagation constant Γ:

$$\Gamma = \frac{\omega \int\limits_{S_\perp} \left(\widetilde{\vec{E}} \varepsilon \vec{E} + \widetilde{\vec{H}} \mu \vec{H} \right) ds - j \int\limits_{S_\perp} \left(\widetilde{\vec{H}} \operatorname{rot}_\perp \vec{E} - \widetilde{\vec{E}} \operatorname{rot}_\perp \vec{H} \right) ds}{\int\limits_{S_\perp} \left\{ [\vec{E}, \widetilde{\vec{H}}] + [\widetilde{\vec{E}}, \vec{H}] \right\} \vec{ds}}. \tag{10}$$

Here $S_\perp$ is the transverse section of the waveguide, and the same symbol ($\perp$) at the differential operator denotes that we put $\partial/\partial z = 0$.

REFERENCES

1. A.D. Berk. Variational principles for electromagnetic resonators and waveguides, IRE Trans., 1956, AP-4, 2, 104.
2. W. Hauser. On the theory of anisotropic obstacles in cavities, Quart. J. Mech. and Appl. Math., 1958, 11, 1, 112.
3. W. Hauser. Variational principles for electromagnetic wave guides in anisotropic materials, Quart. Appl. Math., 1958, 16, 3, 260.
4. V.V. Nikol'skiy. Investigations of hollow systems with anisotropic regions by the eigenfunction method. Part I. Resonator, Radiotekhnika i elektronika, 1960, 5, 11, 1802.
5. V.V. Nikol'skiy. Investigation of hollow systems with anisotropic regions by the eigenfunction method. Part III. Waveguide, Radiotekhnika i elektronika, 1961, 6, 1, 74.
6. P.M. Morse and H. Feshback. Methods of Theoretical Physics, McGraw-Hill, New York, 1953, v. II.

Received by editor 9 December 1960

---

*The quantities $\widetilde{\vec{E}}$ and $\widetilde{\vec{D}}$ in formulas (8a) and (8b) must be regarded as tending to the eigenfunctions of the operators related with $\frac{1}{\epsilon}$ rot $\frac{1}{\mu}$ and rot $\frac{1}{\mu}$ rot $\frac{1}{\epsilon}$ respectively.

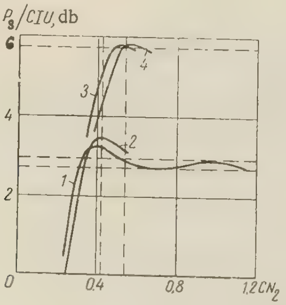
EFFECT OF SECTIONALIZATION ON THE SATURATION POWER OF A TRAVELING WAVE TUBE

A.I. Denisov

As is well known, one of the methods of ensuring stability of a traveling wave tube is to sectionalize the slow-wave system by introducing a drift portion [1, 2, 3].

We consider below the effect of the position of the discontinuity (a very short drift) on the saturation power of a tube for the case when both sections have fully matched loads.

Figs. 1 and 2 show the results of integration of the nonlinear equations [4] of a TWT with two sections, as carried out with the "Kiev" computer.



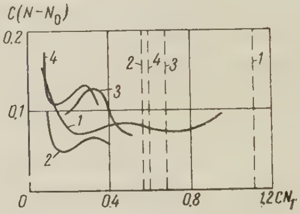


Fig. 1. Dependence of saturation power in a traveling wave tube with break on the length of the output section.

Fig. 2. Effect of sectionalization on the length of the tube.

At the point of discontinuity, the variables characterizing the motion of the electrons are continuous, and the power flux at the output of the gathering section is zero.

What is common to both figures are the following parameters: (we use Pierce's notation [1] throughout): the amplification parameter $C = 0.05$, the space-charge parameter $GC = 0$, the damping parameter $d = 0.1$.

Fig. 1 shows the dependence of the relative saturation power in a TWT with discontinuity, P_S/CIU in decibels (I — current, U — voltage of the electron beam) on the relative length of the output gathering section CN_2 (it is assumed that the output of the gathering section coincides with the maximum-power point). Curves 1 and 2 correspond to an electron velocity parameter $b = 0$ and input power levels P_{in}/CIU of -32 and -6 db; curves 3 and 4 pertain to $b = 1$ and to input powers of -10.4 and -6.4 db (the parameter $b = 0$ corresponds to the maximum of linear amplification coefficient, and $b = 1$ corresponds to a maximum efficiency in the tube without breakup).

The horizontal dotted lines denote the corresponding saturation-power levels in the TWT without break.

It follows from Fig. 1 that the break begins to reduce greatly the saturation power, if the length of the gathering section CN_2 is less than 0.4 for $b = 0$ and less than 0.54 for $b = 1$. As the length of the gathering section CN_2 is reduced from 0.42 to 0.22 (curve 1 of Fig. 1), the saturation power drops by 3.2 db.

Fig. 2 shows the difference $C(N - N_0)$ between the length of a tube with break, $CN = CN_1 + CN_2$ (from the input to the maximum power) and the length of a tube without break for identical parameters b and P_{in}/CIU.

The abscissas represent the length of the input bunching section CN_1. The numbering and the parameters of the curves on Fig. 2 are the same as in Fig. 1. The vertical dashed lines denote the corresponding lengths CN_0 of a TWT without break.

As follows from Figs. 1 and 2, when the break approaches the input to the tube, the saturation power remains almost constant; but if the length of the bunching section becomes small ($CN_1 \simeq 0.18$), then the length of the tube CN increases sharply as the other parameters remain fixed.

Thus, when $b = 1$ it is desirable to have the length of the input and output sections not less than $CN_1 = 0.18$ and $CN_2 = 0.54$ respectively. A tube with such sections has a gain of about 12 db in the saturation mode with $b = 1$.

It must be noted that in a real traveling-wave tube the influence of the space-charge forces on the drift portion may be appreciable.

REFERENCES

1. J.R. Pierce. Traveling-Wave Tubes, Van Nostrand, N.Y. 1950, Chapter IX.
2. J.D. Pearson, H.S. Cockroft. A 20 kw pulsed traveling-wave tube. Proc. IRE, 1958, p. B, 105, Suppl. No. 10, 458.
3. A.I. Denisov. Calculation of gain of sectionalized traveling wave tube. News of the Universities, Ministry of Higher Education USSR — Radio Engineering, 1961, 4, 2, 212.
4. G.N. Rapoport. Nonlinear theory of traveling-wave tube (review), ibid, 1958, 1, 5, 599.

Received by editors 23 March 1961

RADIO SPECTROSCOPE WITH DISK RESONATOR

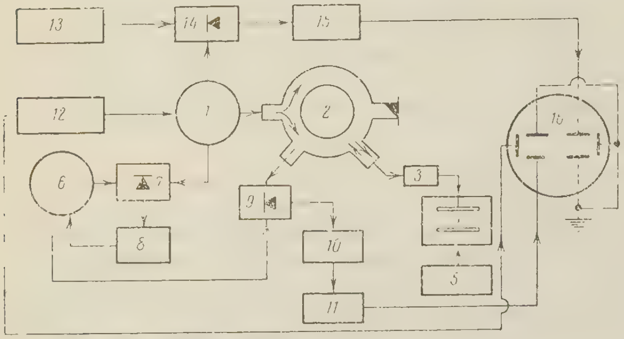
T.M. Murina

Radio spectroscopes with molecular beams, in which the molecules are sorted by states [1], have recently assumed an important role with respect to increasing the resolving power of spectroscopes. Radio spectroscopic means permit investigation of the hyperfine structure of molecules in beams, with a high degree of accuracy. However, the resonators used do not make Stark modulation possible, nor do they permit operation in the range from centimeter to submillimeter wavelengths. Since the high frequency field in a disk resonator is concentrated inside the resonator, the Stark lines do not become broadened by edge effects. Consequently, it is possible to measure in such spectroscopes the dipole moments with high degree of accuracy. The construction of a disk resonator makes it possible to use several beams simultaneously, so that the sensitivity is increased.

The main task in our investigation was the development of a superheterodyne radio spectroscope of the resonator type with Stark modulation. We propose henceforth to work with beams.

A block diagram of the spectroscope is shown in the figure. We used in the installation a disk resonator (2) of the reflex type with a ferrite circulator. The use of such a resonator simplifies the tuning of the resonator frequency. The resonator was made of brass, and the disk surfaces were lapped but not polished. The Q of such a resonator is on the order of 2,000 — 3,000. We did not attempt to increase the Q of the resonator, so as to permit observation of a relatively broad frequency spectrum in the resonator band. The mode excited in such a resonator corresponds to a plane wave, and the electric vector of the microwave field is parallel to the planes of the disks. Since an ordinary spectroscope with cavity resonator has a small volume and high Q, saturation sets in at very low powers. The disk resonator has a much greater volume than the ordinary resonator, and consequently saturation sets in at greater power levels but still lower than in a waveguide cell. When the power level is decreased by saturation, the sensitivity of the video spectroscope decreases, and it becomes necessary to employ superheterodyne reception.

The heterodyne used in our setup was a klystron operating 27 Mc away from the frequency of the signal klystron. The frequency difference between the signal and heterodyne klystrons was maintained constant by phase stabilization of the heterodyne klystron against the signal klystron. In our system the signal-klystron frequency could be tied in with the harmonic of a quartz-controlled generator. This was not done in our experiment, since high stability of the signal-klystron frequency was not required. The signal klystron frequency was made to swing over several megacycles by a 50-cps sawtooth voltage. A square-wave 75-kc voltage from the Stark modulator was applied to the disks, one of which was grounded.



Block diagram of spectroscope:

1 — Signal klystron; 2 — ferrite circulator; 3 — wave meter; 4 — disk resonator; 5 — Stark modulator; 6 — heterodyne klystron; 7 — mixer head; 8 — phase-stabilization block; 9 — mixer head; 10 — IF amplifier, 27 Mc; 11 — low-frequency amplifier, 75 kc; 12 — sawtooth generator; 13 — quartz frequency multiplier; 14 — mixer head; 15 — KV-M receiver; 16 — double-beam oscilloscope.

The absorption-line frequency was measured by means of frequency markers obtained from harmonics of a quartz-controlled multiplier and the output signal of the klystron. The absorption line and the frequency markers were viewed simultaneously on the screen of a double-beam oscilloscope. The operation of the setup was tested at 15,000 Mc and the dipole moment of the formaldehyde molecule (CH_2O) was measured. Measurement of the dipole moment in a Stark spectroscope with a waveguide cell always entails many shortcomings, which introduce a large error in the measurement of the dipole moments. The use of a disk resonator enables us to measure with great precision the distance between disks, and consequently the intensity of the electric field, and also to exclude the need for calibrating the cell against molecules whose dipole moment is known with great accuracy. Whereas in waveguide cells one observes the σ-Stark components corresponding to the case of parallel fields, in a disk resonator one observes the π-Stark components corresponding to the case of perpendicular fields. A determination of the dipole moment of formaldehyde was made by measuring the difference in frequency between the Stark lines and the undisplaced line in the $2_{11} \rightarrow 2_{12}$ transition. The observed structure coincided fully with the calculated one. The dipole moment was found to be $\mu_0 = 2.34$ Debye units. The accuracy of the measurement did not exceed 1%, owing to technical factors. This value of the dipole moment coincides with that obtained in reference [3].

REFERENCES

1. Quantum Electronics, N.Y., 1960
2. A.I. Barchukov and A.M. Prokhorov. Radiotekhnika i elektronika, 1959, 4, 12, 2094.
3. J. Phys. Soc. Japan, 1960, 15, 2.

P.N. Lebedev Physics Institute
Academy of Sciences USSR

Received by editors 13 February 1961

ELECTRON FOCUSING IN THE CATHODE SYSTEM OF PHOTOMULTIPLIERS WITH SEMITRANSPARENT CATHODES

A. G. Berkovskiy and Chia Kao

Photomultipliers with semitransparent cathodes, deposited on the internal end surface of a glass bulb, are used extensively in the investigation of weak scattered light. In the simplest and presently most widely used version the electron-optical cathode system of a multiplier with end cathodes is made up (Fig. 1) of the cathode surface a, a metallic layer (collar) having the same potential as the cathode sputtered on the side surface b of the bulb adjacent to the cathode, and a diaphragm c on which the potential accelerating the photoelectrons is applied.

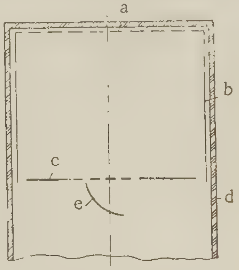


Fig. 1. Cathode electron-optical system of photomultiplier with semitransparent cathodes: a — cathode; b — collar; c — diaphragm; d — bulb; e — first emitter.

Such a system focuses the electrons emitted from the surface of the cathode into the opening of a diaphragm located at the symmetry axis of the system. The area of the opening is several times smaller than the area of the cathode. The fact that the field of the first stages of the multiplier system sags slightly in the input aperture of the diaphragm in the cathode space does not affect the trajectories of the photoelectrons at all, and will therefore be disregarded.

In spite of the very extensive use of this system, no descriptions of its electron-optical properties have been published to date.

Wendt [1], in a theoretical investigation of the wave form of the pulses in the cathode circuit of this system, approximated the electrostatic field by means of the function

$$\varphi = ER \sin h \frac{z}{A} J_0\left(\frac{r}{A}\right),$$

where R — radius of collar; A = 2.408; E — field intensity at the center of the cathode; r and z — cylindrical coordinates; J_0 — Bessel function of zero order.

The use of Wendt's formula for a study of the focusing properties of the system involves very cumbersome computations.

In the system shown in Fig. 1, according to a communication by Bernhardt [2], the ratio of the photoelectron beam diameter in the plane of the diaphragm to the diameter of the cathode fluctuates from 1/3 to 1/8. The author unfortunately did not indicate at what ratio of the cathode diameters to the cathode-diaphragm distance these data were obtained. In addition, it is not clear whether these data were obtained by experiment or by calculation.

The electron-optical (focusing) properties of the system shown in Fig. 1 are characterized, in our opinion, by the two dimensionless parameters [3]

$$p = \frac{D}{d} \text{ and } q = \frac{D}{h},$$

where D — diameter of the collar; d — diameter of photoelectron beam in the plane of the

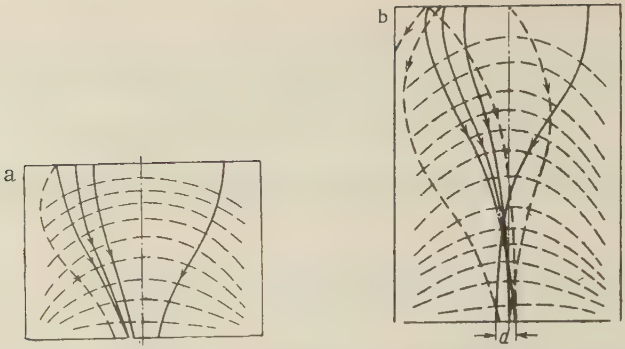


Fig. 2. Results of simulation of the system of Fig. 1 in an electrolytic trough
a — $q = 1.4$; b — $q = 0.8$

diaphragm (diameter of entrance window of the diaphragm); h — distance between the planes of the cathode and the diaphragm.

In the design of a photomultiplier one usually specifies the diameter of the cathode, which is determined by the principal field of application of the new instrument. The diameter of the entrance opening of the diaphragm depends on the chosen design of the multiplier system, which generally speaking is also determined by the end purpose of the multiplier. This fixes the parameter p. The parameter q is obviously uniquely determined by the fixed value of p.

As already noted, an analytic derivation of the formula $p = f(q)$, even without allowance for the initial velocities of the photoelectrons, calls for laborious calculations. It is much simpler to construct this function by simulation (in an electrolytic trough) of a system with different values of q, by determining d (and consequently p) in each version after constructing the trajectories of the photoelectrons emitted from different points of the cathode.

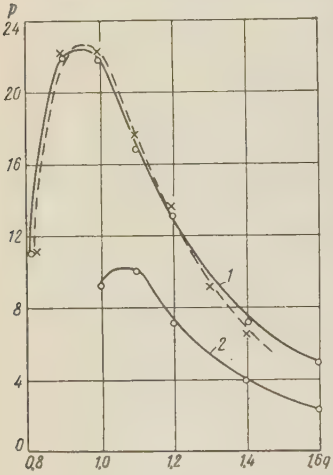


Fig. 3. Curves showing focusing of electrons in the electron-optical cathode system of Fig. 1.

The model we used was a half-cylinder with one fixed end (cathode) and a movable cover (diaphragm). By varying the distance between the "cathode" and the "diaphragm" we could simulate systems with different values of p. The electrodes were made of stainless sheet steel. The internal diameter of the cylindrical surface was 20 cm. The diameter of the working surface of the cathode was usually somewhat smaller than the diameter D of the collar. We chose a ratio $D_c = 0.8$ D, which is frequently encountered in practice.

Two modulated versions of the system under consideration are shown in Fig. 2. Taking into account the symmetry of the system, it is sufficient to carry out the construction on only one side of the axis. Only the extreme trajectory was usually plotted in the second half of the field. The mirror symmetry of the extreme trajectories served as a criterion of the accuracy of the duplication of the potential relief and the correctness of the trajectories.

The function $p = f(q)$ were plotted for two cases (Fig. 3): Curve 1 — the photoelectrons are emitted with zero initial velocities, $v_0 = 0$; Curve 2 — $v_0 = 0.4$ v; the voltage between the cathode and the diaphragm was 200 volts; the electrom emission angle was 45°.

The trajectories were plotted in the near-cathode region by the radius-of-curvature method, while the sine law was used in the central and end regions.

Curve 1 is of interest only in the case of very large field intensities near the cathode, something rarely encountered in practice. The curve has a maximum corresponding to the best focusing of the electrons at $q_{opt} = 0.9 - 1.0$ ($D \simeq h$). If $q > q_{opt}$ ($d < h$), then the cross section of the electron beam with minimum crossover lies beyond the plane of the diaphragm. In the case when $q < q_{opt}$ the crossover is located above the diaphragm. Taking into account the fact that the first emitter on which the photoelectrons must be focused is located some distance behind the diaphragm, one must choose q_{actual} somewhat greater than q_{opt}. The best relation between these two quantities can be determined only for a specific design of the multiplier system.

Curve 1 can be approximated with good accuracy by the empirical formula

$$p = 1 + 1902\,(q - 0{,}7)^{1{,}96}\,e^{7{,}35(q-0{,}7)}.$$

The calculated curve is shown dotted in Fig. 3.

Curve 2 is closer to the real operating conditions of the photomultiplier. For this curve $q_{opt} \simeq 1.1$. When $q < 1$ the field intensity at the emitting surface is som small that the electrons emitted from the edges of the cathode at an angle of 45° strike the collar, i.e., the working diameter of the cathode is smaller than the chosen value.

A comparison of curves 1 and 2 shows that the proper initial velocities which we chose increase the diameter of the photoelectron beam in the plane of the diaphragm by about two times.

REFERENCES

1. G. Wendt. Ann. radioelectr., 1955, 10, 42, 377.
2. W. Hartmann and F. Bernhard. Fotovervielfacher und ihre Anwendung in der Kernphysik, Berlin, 1957.
3. A.G. Berkovskiy. Dissertation, Moscow Power Institute, 1959.

Received by editor 28 January 1961

SPECTROSCOPIC DETERMINATION OF THE DEIONIZATION RATE OF CESIUM VAPOR IN A MAGNETIC FIELD

Yu. M. Aleskovskiy and V. L. Granovskiy

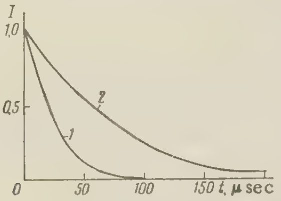
It was established earlier that a homogeneous longitudinal magnetic field slows down the rate of diffusion decay of a low-pressure plasma. It was also found [5] that in the deionization of rarefied helium some of the charges are neutralized by volume processes. No direct observation was made of the volume recombination in a plasma decaying in a magnetic field.

In the present paper we aim to determine directly the effect of the magnetic field on the recombination loss of charged particles in the volume of a decaying plasma. For this purpose we measured the recombination radiation of a cesium plasma in the region of an end-point continuum with a limit 4940 Å, corresponding to the capture of electrons by the cesium ions at the level $Cs6P_{1/2}$ [6]. The intensity of the recombination radiation of the plasma (I) in an electrically positive gas enables us to judge the concentration of the electrons and ions (n) in the plasma, inasmuch as $I \sim n^2$. On the other hand, I indicates how many electrons and ions vanish as a result of radiation combination, since $I \sim (dn/dt)_R$. Oscillograms of I obtained in a decaying plasma enable us to judge the rate of deionization of the gas. It must be kept in mind here that if $n \sim \exp(-t/\tau)$ then $I \sim \exp(2t/\tau)$, where τ is the deionization time constant. Consequently, the reduction in the concentration of the charged particles is twice as slow as the decrease in intensity of the recombination glow.

The cesium plasma was produced by an arc discharge in a tube 2.5 cm in diameter and

40 cm long, fed with half-wave current at line frequency. A thyratron connected in parallel with the tube was ignited at the phase corresponding to the maximum current, and the discharge in the tube was terminated (for a description of this method of obtaining a decaying plasma see, for example, reference [7]). The tube was situated in a solenoid producing a homogeneous magnetic field of intensity up to 1300 oersteds. The plasma radiation was focused on the entrance slit of a monochromator, and a photomultiplier was placed at the monochromator exit. The signal of the photomultiplier was recorded with an oscillograph; the slaved sweep of the oscillograph was synchronized with the instant of termination of the discharge in the tube, and consequently, with the start of the deionization of the gas.

Application of a magnetic field on the plasma greatly increases the intensity of the recombination radiation, owing to the increase in concentration of charged particles in the discharge [8]. In order for the decay of the plasma to begin at equal concentration of the charge particles, the initial intensity of radiation was equalized for $H = 0$ with the intensity for $H \neq 0$ by varying the discharge current.



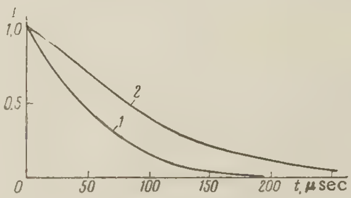


Fig. 1. Intensity of recombination radiation as a function of the time at $p = 5$ μ Hg:
1 — $H = 0$; 2 — $H = 1300$ oersteds

Fig. 2. Intensity of recombination radiation as a function of the time for $p = 15$ μ Hg:
1 — $H = 0$; 2 — $H = 1300$ oersteds

Figures 1 and 2 show the variation of the intensity I of the recombination radiation during the ionization of the cesium vapor. The magnetic field greatly slows down the decay of the plasma, and, as expected, the influence of the magnetic field decreases with increasing pressure. Table 1 shows the values of the time τ necessary to reduce the concentration of charged particles by a factor e, calculated from the deionization curves, for $H = 1300$ oersted.

Inasmuch as $I \sim (dn/dt)_R$, the total number of particles which recombine in this manner within a time $\Delta t = t_2 - t_1$, $N_R = k \int_{t_1}^{t_2} I dt$. The value of kN_R can be determined, accurate to a constant factor, by measuring the area under the curve plotted in I-t coordinates (Figs. 1 and 2). Table 2 lists the ratios $N_R(H)/N_R(0)$ of the number of particles that disappear as a result of radiation recombination at $H = 1300$ oersted and $H = 0$. It is seen from Table 2 that at low pressures the magnetic field increases noticeably the fraction of charged particles which recombine with radiation in the volume of the gas.

Table 1

p, μ Hg	$\tau(0)$, μ sec	$\tau(H)$, μ sec	$\tau(H)/\tau(0)$
7	48	160	3.3
10	62	180	2.9
15	94	270	2.88
25	108	270	2.5
50	168	280	1.67

Table 2

p, μ Hg	$N_R(H)/N_R(0)$
	$\Delta t = 0 - 250$ μ sec
7	2.2
10	1.9
15	1.6
25	1.5
50	1.1

Thus, it is shown by direct spectroscopic observations that the magnetic field increases the average lifetime of the charged particles and slows down the diffusion decay of the plasma, contributing thereby to the recombination charge losses in the volume of the gas.

REFERENCES

1. W. Bostik, M. Levine. Phys. Rev., 1959, 97, 13.
2. A.S. Syrgiy, V.L. Granovskiy. Radiotekhnika i electronika, 1960, 5, 7, 1129.
3. S.K. Guzhova, A.S. Syrgiy. Radiotekhnika i elektronika, 1960, 5, 9, 1516.
4. V.E. Golant, A.P. Zhilinskiy. ZhTF, (Journal of Technical Physics), 1960, 30, 7, 745.
5. A.S. Syrgii, V.L. Granovskii. Radiotekhnika i elektronika, 1960, 5, 9, 1522.
6. F. Mohler, C. Boeckner. J. Res. Nat. Bur. Standards, 1929, 2, 489.
7. V.L. Granovskiy. News of the Academy of Sciences USSR, Physics Series, 1938, 4, 419.
8. Yu. M. Alekseyevskiy, V.L. Granovskiy, E. Mikhalets. Radiotekhnika i elektronika 1961, 6, 4, 674.

Physics Faculty,
M.V. Lomonosov Moscow State University,
Chair of Electronics

Received by editor 18 April 1961

RADIO ENGINEERING AND ELECTRONIC PHYSICS

Institute of Radio Engineering and Electronic Physics,
Academy of Sciences of the USSR

EDITORIAL BOARD

Editor-in-Chief: V.A. Kotel'nikov
Associate Editors: D.V. Zernov, Yu.B. Kobzarev

A.I. Berg
B.A. Vvedenskiy
I.S. Gonorovskiy
V.L. Granovskiy
L.A. Zhekulin
N.D. Devyatkov

L.N. Dobretsov
A.N. Kazantsev
S.G. Kalashnikov
P.L. Kapitsa
V.V. Migulin
A.L. Mikaelyan
A.A. Pistol'kors

A.M. Prokhorov
S.M. Rytov
V.I. Siforov
Ya.N. Fel'd
S.E. Khaykin
B.M. Tsarev

Scientific Secretary of Editorial Board: G.A. Bernashevskiy

The English Edition of Radio Engineering and Electronic Physics is mailed to subscribers within 18 weeks after the publication of the original Russian issue.

Russian electronic journals published by the
American Institute of Electrical Engineers
Translated by Royer and Roger, Inc.

Subscription rates

	Individuals		Libraries	
	$	£	$	£
Radio Engineering and Electronic Physics	28.50	10	57.00	20
Radio Engineering	14.25	5	28.50	10
Telecommunications	14.25	5	28.50	10

Royer and Roger translates and produces
the following Russian scientific journals:

Biophysics
Entomological Review
Geochemistry
Geodesy and Cartography
Izvestiya, Academy of Sciences of the USSR, Geologic Series
Pavlov Journal of Higher Nervous Activity
Problems of Oncology
Radio Engineering
Radio Engineering and Electronic Physics
Refractories
Sechenov Physiological Journal of the USSR
Soil Science
Telecommunications

Comments and inquiries regarding *Radio Engineering and Electronic Physics* and other translation journals should be sent to:

International Division
Royer and Roger, Inc.

1000 Vermont Avenue, N.W.
Washington 5, D.C.

41 East 28th Street
New York 16, New York

RADIO ENGINEERING AND ELECTRONIC PHYSICS

РАДИОТЕХНИКА И ЭЛЕКТРОНИКА

NUMBER 9 1961

CONTENTS

Yu.B. Kobzarev, A.Ye.Basharinov: Effectiveness of Search Algorithms Based on the Use of Trial Steps of Controlled Duration 1253
V.V. Vitkevich, A.D. Kuz'min, et al.: Radio Astronomic Observations of Soviet Space Rockets 1261
N.D. Filipp: Character of Fluctuations of a UHF Radio Signal Propagating over an Inhomogeneous Surface 1270
N.K. Kul'man, R.L. Stratonovich: Certain Optimal Devices for Detection of a Pulse Signal of Random Duration in the Presence of Noise 1279
V.V. Shirokov: Effect of Signal Fluctuations on a Receiver with AGC 1288
V.K. Turkin, G.A. Levin: On the Detection of Frequency Modulated Oscillations . . 1296
G.A. Levin, V.K. Turkin: On the Theory of Detection of Complex Frequency Modulated Oscillations 1301
V.A. Shpolyanskiy, L.S. Tyufyakin, P.P. Korsakov: Automatic Phase Control with Delay 1304
B.M. Minkovich, Yu.I. Davidchevskiy: Analysis of Antennas with Plane Apertures . . 1317
L.A. Vaynshteyn G.D. Malyuzhinets: Transverse Diffusion in Diffraction by an Impedance Cylinder of Large Radius. II. Asymptotic Laws of Diffraction in Polar Coordinates 1324
B.N. Morozov: Influence of Bending on the Properties of Periodically Diaphragmed Waveguides of Rectangular Section 1331
Yu.N. Pchel'nikov: Interaction between an Electron Beam and the Field of a Helix in the Region of an Absorbing Insert 1335
G.F. Filimonov: Excitation of a Delay Line by Means of a Periodic Sequence of Ideally Focused Electron Bunches 1341
V.N.Shevchik, A.A. Kurayev, General Dispersion Equation of a Traveling Wave Tube with Periodic Slow-Wave System 1351
A.M. Starik: Propagation of Microwave Oscillations in a Waveguide Containing the Cathode Parts of a Glow Discharge 1365
T.F. Loginova: Hollow Electron Beam of Conical Shape in a "Conical" Magnetic Field 1372
G.S. Misezhnikov, V.B. Shteynshleyger: Contribution to the Theory of a Traveling-Wave Quantum Paramagnetic Amplifier 1377
A.S. Maydankovskiy: Investigation of a Transistorized Low-Frequency Self-Oscillator 1385
A.M. Strashkevich: Investigation of Non-Gaussian Trajectories in the Field of an Electrostatic Quadrupole Lens 1392
V.N. Lutskiy: Procedure for Energy Analysis of Thermionic Electrons under Pulsed Emission Conditions 1396

BRIEF COMMUNICATIONS

V.I. Bunimovich, V.A. Morozov: Estimate of Frequency of Narrowband Signal against a White-Noise Background 1403
A.V. Prosin: Calculation of Reliability of Tropospheric Communications System in the Presence of Correlated Fading 1407
A.L. Fel'dshteyn: Cascaded Bridges 1411
V.V. Nikol'skiy: Variational Principle for Hollow Systems with Anisotropic Medium 1413
A.I. Denisov: Effect of Sectionalization on the Saturation Power of a Traveling Wave Tube 1416
T.M. Murina: Radio Spectroscope with Disk Resonator 1417
A.G. Berkovskiy, Chia Kao: Electron Focusing in the Cathode System of Photomultipliers with Semitransparent Cathodes 1419
Yu.M. Aleskovskiy, V. L. Granovskiy: Spectroscopic Determination of the Deionization Rate of Cesium Vapor in a Magnetic Field 1421